

War against famine, for nutrition

In Somalia, the world is fighting an unprecedented war in the name of basic nutrition. In the United States, the government has gone overboard on the matter of nutrition and health.

THE world embarked on a perilous military venture last week when the United Nations resolved to invade Somalia for humanitarian purposes. An estimated two million Somalis, their stomachs bloated, their ribs showing, are said to be in danger of imminent starvation because warring clan lords are stealing food from relief agencies.

Faced with the sight of dying women and children, soldiers from as many as 11 nations, led by the US military, will march into Somalia, a desolate, government-less country, to distribute food—nothing more, nothing less. The White House says the purpose of the UN mission is “not to establish a new government or resolve the decades-long conflict there or to set up a protectorate, or anything like that.” Nor is the UN going in to protect broad world interests, as was the case with oil in the Gulf War. Somalia has nothing of strategic significance. It does not even have proper roads.

No high technology

Thus, in a late twentieth century world whose military actions have depended on increasingly sophisticated technology, soldiers are being deployed with no need for ‘smart’ weapons. The sciences of physics, metallurgy and remote sensing apparently have no role to play in a theatre in which the opponents are armed with conventional guns, grenades and potholes. Furthermore, if we are to believe that this UN action is undertaken only to assure the distribution of bags of rice and wheat already in Somali warehouses, there is no call for agricultural science here. There is no green revolution in the cards, just a short-range (and short-sighted) plan to defeat imminent death by distributing bowls of rice. Nutrition at its most basic. In terms of science, this is a throwback to another era. And it is military intervention without a meaningful plan.

There is no little irony then that the day before announcing the despatch of American soldiers to Somalia, US President George Bush made headlines on another (equally low-technology) nutritional front. The president stepped into a bitter battle between the Secretaries of Health and Human Services (HHS) and Agriculture over nutritional labels on food. Conflict over what to say about fat was particularly intense. (The Somalis have no dietary fat; Americans too much.)

The HHS secretary flavoured a label that puts the content of food into the context of a daily ‘average’ diet of 2,000 calories, restricted to 65 grams of fat. The agriculture

secretary (and the meat and dairy interests his department regulates) wanted to limit labels to food content, with no information about percentages of daily diet. In the event, ‘context’ and HHS carried the day. Next month, new food labelling regulations running to 4,000 pages will become law. The bone thrown to agriculture (thank heaven) is an agreement that food labels will not be required on restaurant menus, leaving people to dine in peace.

Dietary guidelines

The label, which will appear in standard format on virtually all foods sold in the United States, is predicated on the hopeful presumption that enough is known about nutrition to establish valid dietary guidelines for a large and diverse population. The premise is open to challenge (as this journal has said before). But even if that were not the case, one look at the new labels is enough to discourage even the most ardent fitness buffs. Standard serving sizes are established as 1/2 cup or 114g. In a sample label, then, total fat is listed as 13g or 20 per cent of the standard daily diet of 2000 calories. In the next line, saturated fat is reported to be 5g or 25 percent of the total diet. The quantities and percentages of cholesterol, sodium, sugars and dietary fiber are also calculated, leaving the consumer in need of a mainframe computer to translate it all into practical terms. Indeed, the information is all but incomprehensible.

Hunger, satiety and longevity are basic drives fuelling the course of politics in two very different but significant ways. Perhaps it is inevitable that the UN decided to authorize military intervention in Somalia. There is a moral imperative to prevent millions of innocent people from dying needlessly. But it is not as simple as Bush stated when he said that famine in Somalia is a tragedy “we could do something about”. Soldiers will lose their lives; the Somalis will not really be helped without a UN commitment to the long run, but the prospect of Western soldiers in Somalia for years is evidence only that the way out of one quagmire (imminent starvation) is to march into another.

On the US domestic front, the plan to put incomprehensible nutritional information on more than 300,000 foods (a bureaucratic nightmare that could well spill over to the European Communities) raises the spectre of a different quagmire. In both cases, last week’s political actions are based on compassion and good intentions, but they are not smart. □



Japan's MITI opens domestic R&D projects to foreign participation

Tokyo. Japan's 'technoglobalism', a policy of the Ministry of International Trade and Industry (MITI) aimed at cultivating international collaboration in research and development, seems to be blossoming. A week ago, the *Nikkei Shimbun*, Japan's leading financial newspaper, revealed that several non-Japanese companies, including the US semiconductor manufacturer Texas Instruments, are likely to join a huge group of Japanese industrial partners in a large-scale MITI-funded project to develop nanotechnology (see below). And foreign participation in other MITI research and development projects is on the increase.

Texas Instruments, Motorola and the Japanese subsidiary of the US chemical company Dupont are reported to be among several foreign companies that have applied to join the \$200-million nanotechnology project along with 39 leading Japanese companies, such as Toshiba, NEC, Hitachi, Fujitsu and Nippon Steel. Participants will not be officially announced until the project starts next month, but Don Shaw, director of Texas Instruments' new research institute in Tsukuba science city, northeast of Tokyo, confirms that his company has applied as one member of a large research association of more than 40 companies.

More than a dozen non-Japanese companies have signed up for various MITI projects in the past two or three years (see table) in response to the ministry's new policy of encouraging foreign participation. Technoglobalism was officially launched by MITI in 1990 at a meeting of the Organization for

Economic Cooperation and Development to counter the perceived 'technonationalism' of the United States, by which MITI means US efforts to protect its domestic technological advances from countries such as

space companies in a large-scale project to develop a turbofan/ramjet for hypersonic transport at three to five times the speed of sound. Winning this foreign participation proved to be a major ordeal for the ministry.

The ten-year project began in 1989, but it was only two years later that the foreign partners, Rolls-Royce of the United Kingdom, SNECMA of France, and United Technologies and Pratt & Whitney of the United States were able to join.

Differences over intellectual property rights were a major obstacle. In the United States and Europe, private companies that join government-sponsored programmes are usually free to exploit the fruits of their research, but in Japan, the government retains control of intellectual property rights. MITI had to rewrite Japanese law before foreign companies could join (see *Nature* **350**, 102; 1991).

Organizational structure also caused some friction. The Japanese companies in the hypersonic project formed a research association and negotiated with MITI as a single group, but US and European companies wanted to deal with MITI individually. MITI established an intermediary committee to act as go-between.

Language difficulties have also made it difficult for the foreign companies to benefit fully from the project, says Tony Millington, president of Rolls-Royce (Far East). Nevertheless, Rolls-Royce and the other non-Japanese companies were eager to join because Japan is one of the few governments putting money into engines for hypersonic transport, which remain much further from commercial development than mere supersonic technology.

The distance from commercial application is also an attraction for non-Japanese companies wanting to join MITI's nanotechnology project. "It's a very long-range project and it makes sense to have international collaboration at this stage" says Texas Instruments' Shaw.

Building on its experience, MITI is now beginning to involve foreign organizations in the very early planning stages of new projects. Next week, several non-Japanese organizations, including the Hydrogen Industry Council of Canada, will participate in a meeting to plan two projects of MITI's New Sunshine Program, a huge programme aimed at developing alternative energy sources and 'environment-friendly' technology. The programme will get under way next fiscal year.

David Swinbanks

Growing foreign participation in Japanese national research

■ Large-scale projects

Super/hypersonic transport propulsion system
Rolls-Royce plc (UK), SNECMA (France), United Technologies Corporation/Pratt & Whitney (USA), General Electric Co. (USA)
Advanced chemical processing technology
SRI International (USA)
Micromachine technology
IS Robotics Corporation (USA), SRI International (USA), Royal Melbourne Institute of Technology (Australia)

■ Basic technologies for future industries

Non-linear photonics materials
BASF AG (Germany)
Molecular assemblies for a functional protein system
GBF (Gesellschaft für Biotechnologische Forschung) (Germany)
High-performance materials for severe environments
Crucible Materials Co. (USA)
New models for software architecture
SRI International (USA)
Production and utilization technology of complex carbohydrates
Pharmacia LKB Biotechnology AB (Sweden)
Quantum functional devices
Motorola Inc. (USA)

Japan. The Japanese are also anxious to forestall criticism that Japan gets a 'free ride' on Western scientific and technological developments.

MITI's first success in drawing foreign companies into a Japanese research and development programme came with the participation of four US and European aero-

Very small plans

Tokyo. MITI's large-scale nanotechnology project, which has attracted the participation of non-Japanese companies (see above), is beginning to take on a more clearly defined form. Called the "Angstrom Technology Project" when announced by MITI last year (see *Nature* **352**, 650; 1991), its title is now "Ultimate Manipulation of Atoms and Molecules". More than ¥25,000 million (US\$200 million) will be invested over ten years.

The project will develop tools, such as the scanning tunnelling microscope, for probing and observing materials at the molecular and atomic level. MITI also hopes to develop technology for reading the structure of DNA and for manipulating atoms and molecules on surfaces and in three-dimensional space. The project will also include development of computer simulation techniques for the theoretical prediction of surface processes, and development of ultra-high-vacuum technology as well as a tunable femto-second laser for manipulating atoms and molecules.

Much of the research is expected to be carried out at a new interdisciplinary research centre being established in Tsukuba science city as part of a reorganization of MITI's Tsukuba research laboratories.

D.S.

Shake-up planned for Galileo as it steals quietly past Earth and on to Jupiter

Washington. After passing Earth at close range for the second time on 8 December, the Galileo spacecraft is now finally on its way to Jupiter, its ultimate destination. Its main antenna remains partially unfurled, severely curtailing its data transmission capacity, but mission planners at the National Aeronautics and Space Administration

low-gain antenna transmits only a paltry 10 bits per second, compared to the 134 kilobits per second that the high-gain antenna should have provided. By strengthening receiving stations on Earth and using data compression techniques, JPL engineers believe they can squeeze one kilobit per second from the low-gain antenna, but that is still less than one per cent of Galileo's planned capability.

If only the low-gain antenna is working when Galileo reaches Jupiter, it will not be possible to transmit data-intensive digital images in real time. Instead, they will have to be tape-recorded on the spacecraft and played back slowly during quiet periods between encounters with Jupiter's moons. Only 3,000–4,000 images would be obtained in a scaled-down mission, out of an originally planned total of 50,000.

Scientists studying the dynamics of the Jovian atmosphere would be especially hard hit, because much of the planned imaging was intended for the compilation of movies showing Jupiter's swirling weather patterns; they would be obliged to select only a few small areas for limited time-sequence photography. Similarly, 'fields and particles' instruments designed to map Jupiter's magnetic field and its interaction with the solar wind would yield data with much less spatial resolution.

Studies of Jupiter's satellites would fare better in a scaled-down mission. The number of images available would be smaller, but it should still be possible, as Galileo visits different moons, to obtain all the highest priority images. Project scientist Torrence Johnson points out that most of the important scientific information contained in the Voyager images came from a few per cent of the 100,000 pictures taken during the mission, and says that with Galileo it will be possible to predict which are the most essential images. "We are not just going in and wildly shooting frames, hoping to find something interesting", he says.

Unaffected in even a low-gain mission would be the science returned by Galileo's probe, which will split off from the main spacecraft months before the encounter and plunge into Jupiter's atmosphere in

December 1995, as Galileo enters Jovian orbit. Data from the probe will be relayed complete back to the orbiter, stored on tape, and transmitted to Earth at leisure. The probe will obtain the first *in situ* measurements from the atmosphere of a giant gas planet, and because these data are so valuable Johnson suggests that even in a scaled-down mission, without full movie coverage, 80 per cent of atmospheric science objectives can still be met.

Although the Galileo team is hopeful that the high-gain antenna can be brought into full service, Johnson isn't counting on it. He says that if the mission has to be scaled down, project scientists and engineers will quickly start looking for ways to regain as much of the original mission's capability as possible. "My gut feeling is that when people start thinking about the tools available to us," he says, "they will get very innovative."

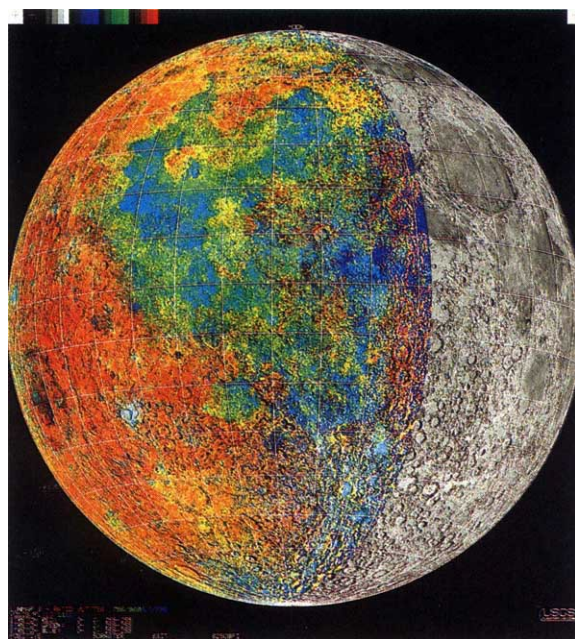
Tony Reichhardt

New human gene therapy institute in Pennsylvania

Washington. Biologist James Wilson, whose gene therapy programme for cystic fibrosis gained government approval last week, will move to the University of Pennsylvania next March to direct a new Institute for Human Gene Therapy. Institute scientists will work on gene therapies for cancer and viral diseases as well as for genetic diseases, which have been the focus of Wilson's research so far. By the end of the century, the Institute is expected to employ 25 researchers and occupy 40,000 square feet of space.

The gene therapy programme Wilson leaves behind at the University of Michigan, in which he treated several patients with low-density lipoprotein receptor deficiency, is similar to the work he will direct at the Pennsylvania institute. The resemblance is not accidental: the Michigan programme was founded by William Kelley, who was chair of Michigan's department of internal medicine and a graduate advisor to Wilson, and is now dean of Pennsylvania's School of Medicine. Kelley seems keen to duplicate Michigan's thriving gene therapy programme at Pennsylvania, which has few gene therapy researchers outside the School of Veterinary Medicine. At Pennsylvania, Wilson plans to recruit researchers and establish ties to the biomedical industry, but his first priority is to start treating cystic fibrosis patients.

Traci Watson



The Moon, pictured by Galileo during this week's fly-by. The colours represent surface brightness in different wavelength bands, from infra-red to ultraviolet, and can be interpreted in terms of surface composition. Green and yellow indicate higher concentrations of iron and magnesium.

(NASA) are still confident that the stuck hardware can be jiggled loose by a series of manoeuvres scheduled to begin later this month.

Engineers at the Jet Propulsion Laboratory (JPL) in Pasadena, California, believe the umbrella-like antenna failed to open fully because three of its 18 ribs are caught against a supporting tower. The spacecraft has been repeatedly turned on its axis, taking the antenna in and out of direct sunlight, in the hope that thermal stress would release the ribs. But this trick has failed, and JPL engineers now have a more aggressive plan. Starting on 29 December, they will send rapid pulses to a drive motor, hoping to pop the ribs loose.

This will be a last-ditch effort, and project managers have already decided that if it fails they will begin to design a scaled-down mission using only a smaller, low-gain antenna to transmit data from Jupiter. The

The young and the restless

Madrid. Hundreds of Spanish postdoctoral fellows, sent abroad for scientific training as Spain's economy took off in the 1980s, now find themselves stranded as the country's finances have faltered and the jobs to which they were expecting to return have failed to materialize. A new initiative by the cash-strapped government to provide short-term contracts for returning researchers has been decried as too little, too late, especially as the number of permanent positions is shrinking.

In 1984, post-Franco Spain began a concerted attempt to raise the low level of its research base, leading to the formulation in 1986 of the national research and development plan, which laid out a framework for research and development along European lines. As part of the plan, young scientists were sent abroad for pre- and postdoctoral training with an expectation of secure employment on their return to Spain. But the high hopes crashed along with the economy, and most of the new-minted researchers found themselves with little hope of a permanent job. Some short-term contracts, with poor employment conditions, were available, but from 1985 only 200 such contracts a year were offered — not enough to go round.

Moreover, Spain started and is continuing a campaign to train more PhDs at home: in 1991, more than 10,000 graduate fellowships were provided. According to José Mato, president of Spain's Council for Scientific Research (CSIC), the idea was to secure Spain's future by substantially increasing the research population. "Although it sounds obvious", he says, "to develop science in Spain, firstly we had to have scientists, and secondly we had to create for them places to work."

The second part of the plan has not been as successful as the first. Angel Pestaña, a molecular biologist at the CSIC Institute of Biomedical Research in Madrid, says that after a brief rise in the number of permanent positions for young scientific investigators in the early 1980s, the ratio of short-term to permanent contracts for young scientists has risen dramatically. The ratio in 1991 was about 4:1, but in Pestaña's discipline it is now about 10:1.

Roberto Fernandez de Caley, director general of the Ministry of Education and Science, says that the government pays travel expenses for the Spanish researchers to come home, but cannot afford to give them permanent jobs. He estimates that about two-thirds find temporary employment in research institutes and universities, but exact statistics are hard to come by because the government does not keep track of where the returning researchers end up.

Young scientists who have gone abroad are at a disadvantage because they must

compete for jobs from a distance and against an increasing number of homegrown researchers. Permanent university jobs are advertised only in a closed circulation university magazine. Many returning scientists therefore take up, as an interim measure, special contracts which originally covered wages only and gave none of the usual employment rights, such as pension benefit or social security. Researchers on these contracts work on established projects and have no chance to develop their own research interests.

In response to the general dissatisfaction with this system, the government has established this year a new system of three-year contracts with full employment benefits. And there are more of them — 300 for 1992 and the same for next year. Most researchers, however, find this an insufficient improvement, particularly because it does not address the more basic problem of a shrinking availability of permanent research jobs in both the universities and industry.

The government is sensitive to the problem, but with its restricted budget it has no easy solution. In addition to tinkering with its short-term contract options, it is now looking to outside sources to help it out. Scientists are being encouraged to apply for funding from the European Community, and Spanish industry is being encouraged to share the burden (see below).

Not everyone is sympathetic to the plight of researchers. Iñigo Aguirre de Cárcer, subdirector at the Ministry of Education and Science, says that researchers make life harder for themselves because they are generally unwilling to consider positions outside the main centres in Madrid and Barcelona. Spain has a policy of devolving research and some openings are appearing in the provinces. But the virtual absence of scientific support in the provinces, compared to Madrid, which hosts at least half of all Spain's scientific activity, is unattractive for the ambitious.

Alison Abbott

Running out of steam ...

Munich. Since the restoration of democracy in the late 1970s, Spanish research has grown much faster than in other Western economies. In the past ten years, Spain has more than doubled the percentage of gross national product (GNP) it allocates to research. But now the government says that the rate of growth cannot be sustained, and science is likely to stagnate at a level that remains one of the lowest in Europe.

In 1983, Spain put only 0.48 per cent of GNP into research, much less than the 2 per cent average among Europe's richer nations. Since then, the government has raised research budgets generously — an average annual increase of more than 15 per cent — and by last year research spending was up to 0.9 per cent of GNP. But economic collapse

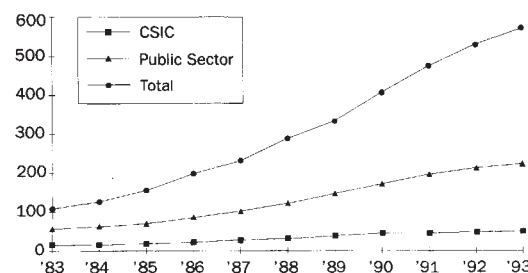
has forced the government to think again about the speed of modernization. In 1992 the science budget grew at a rate barely above inflation, and something similar is expected for 1993.

The poor economy has necessitated hard thinking about spending priorities, but has Spain also lost the political will to bring its research base into line with the European Communities (EC)? 'Not so', says Elias Fereres, secretary of state for universities and research. Some decisions demonstrate the government's continuing commitment to research, he says; for example, education and science will be the only ministry exempt from the freeze on civil service appointments announced for 1993.

Fereres also stresses the positive achievements of Spanish research before the brakes were put on growth. Scientific productivity, as measured by the US Institute for Scientific Information, has doubled along with investment. Spain contributed 1.7 per cent of the world's scientific output in 1991, compared with 0.8 per cent in 1984. The quality of papers has also increased, as judged by the usual parameters of impact factors and publication in English-language, refereed journals. Measured per researcher, Spanish output is on a par with the richest countries. And the introduction of the national plan in 1986, which for the first time forced a logical

Research and development outlays in Spain

Figures in thousand millions*



* 1991–1993 figures estimated

Source: 1983–90: OECD, INE. 1991–93: Nat. Plan, SDIC

structure on its research plans, ensured that Spain established itself in important competitive areas such as biotechnology, engineering and microelectronics.

The increase in Spain's scientific stature has been largely led by the CSIC and its 96 research institutes, 42 of which are in Madrid. But CSIC is now also suffering a budget squeeze, with only a six per cent increase, close to the inflation rate, for 1993. Between 1983 and 1990, the average budget growth was around 17 per cent. There are no strong indications that the universities or industry will pick up the slack. According to Francisco Mora, a university professor in Madrid, "the scientific infrastructure in the universities is far behind that of CSIC". And Fereres admits that despite his government's encouragement, "Spain traditionally imports technology and industry lacks the commitment to do research." In 1988, the government introduced a scheme to pay for one-year contracts for young scientists to work in an industrial environment, and many research-linked tax incentives are now available. But the hope that this would stimulate the fledgling research and development culture in Spanish companies has not yet been fulfilled.

Another source of new money for Spanish science is the EC, and about 1,350 Spanish groups now participate in 955 different EC projects. But they are leaders in only 4 per cent of those projects, which is a measure of how far Spain is away from what Luis Oro, secretary general of the national plan, calls "the ultimate goal of reducing the gap separating Spain from the most advanced industrial countries." **Oscar G. Segurado**

Physics centre threatened

Washington. Fusion research at the Princeton Plasma Physics Laboratory (PPPL), one of the principal US fusion research centres, would be shut down for as long as nine years under a recent plan submitted to the US Department of Energy (DOE). The Fusion Energy Advisory Committee, which was asked by DOE to rank US fusion projects, recommended closure of the Princeton Beta Experiment-Modification (PBX-M) in 1994, on the grounds that the \$10 million annual cost would be better spent on the International Thermonuclear Experimental Reactor, to which the United States is a subscriber, and on PPPL's new project, the Tokamak Physics Experiment. The TFTR (Tokamak Fusion Test Reactor) at Princeton is scheduled to be shut down in 1994, and the Tokamak Physics Experiment is not due for completion until 2003. Closure of PBX-M will displace 25 scientists, unless the budget for fusion energy research rises by 5 per cent for the next several years, which seems unlikely in the light of this year's increase of only 0.8 per cent. **Traci Watson**

Proposal opens debate over US data on animals

Washington. A proposal that the US government should collect more data on research animals has renewed debate about the flaws in the existing system. Although many researchers believe that additional information is unnecessary and could be used by activists campaigning to eliminate animals from the laboratory, most agree that changes are needed in what is now collected.

At present, institutions must submit an annual report to the US Department of Agriculture (USDA) on animals used for

6 per cent of research animals are subject to painful procedures undrugged; many researchers think the figure is too high, and animal advocates contend that it is too low.

The Humane Society of the United States has submitted a proposal to USDA for a pain scale that provides "a better profile of what laboratory animals experience". Their scale has five categories, based on the degree of pain or distress a procedure causes (see figure). Pain scales more extensive than USDA's are already used in some European countries and at least a dozen large US research institutions.

In 1987, USDA proposed a detailed pain scale for use by the committees that oversee animal care at each institution, but it was withdrawn after researchers complained that it was vague and too subjective. Many are also concerned that such information would be used by animal advocates to turn public sentiment against the scientific community.

It is unlikely that USDA will support the Humane Society's request, but there is room for improvement. Many researchers believe that a thorough explanation of the reporting requirements would yield more accurate information, and they would like to work with the agency on such changes. "USDA has never come to the community asking for help", says one researcher.

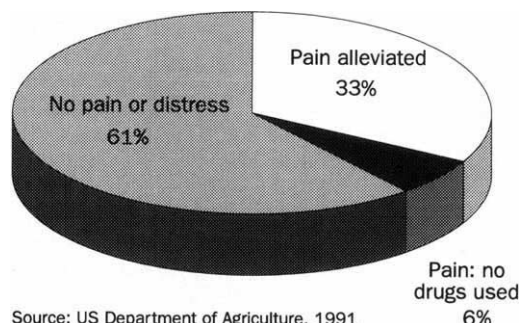
"Doing so would show that they're serious."

Scientists believe they have an ally in Dale Schwindaman, the new director of the USDA department that collects the annual reports. They are encouraged by his agreement to rescind changes in a rule that would have made the statistics less accurate and by his appearance earlier this month at the annual meeting of the American Association of Laboratory Animal Science.

But substantive changes will demand more than just good will. The USDA department that oversees animal research is understaffed and overworked; compiling more information and explaining the current system may be asking too much. Both animal activists and scientists agree that a necessary first step would be a larger USDA budget. **Traci Watson**

Animals in the laboratory

There were 1.8 million animals used in US experiments in 1991, excluding rats, mice and birds.



Source: US Department of Agriculture, 1991

Proposed categories of pain and suffering

Category I: Procedures involving the use of microorganisms, plants, invertebrates, vertebrate embryos, *in vitro* cultures, or no living organisms at all.

Category II: Procedures using vertebrates species that should produce little or no pain or discomfort.

Category III: Procedures that should produce minor or short-term pain or distress in vertebrate animals.

Category IV: Procedures involving significant (moderate to severe) but unavoidable pain or distress to vertebrate animals.

Category V: Procedures involving severe pain near, at, or above the pain tolerance or unanesthetized, conscious animals.

research, testing or education. Scientists have long complained that the rules for the annual report are vague and unreasonable, especially the requirement that all animals be placed into one of three categories that measure pain and distress: those experiencing no pain, those subjected to painful procedures and treated with alleviating drugs and those that undergo painful procedures without any alleviating drugs.

Reporting varies widely between institutions and among investigators. Even a simple injection may or may not be reported as painful, says Franklin Loew, dean of the school of veterinary medicine at Tufts University in North Grafton, Massachusetts.

Researchers and animal advocates agree that the present statistics are probably misleading. For example, USDA says that only

Physics and family values

Quebec. In an attempt to make amends for the publication, in the *Canadian Journal of Physics*, of a work of amateur sociology that blamed feminism and working mothers for the moral failings of modern undergraduates, the National Research Council (NRC) of Canada is taking the unprecedented step of publishing a monograph of rebuttals to the thesis.

The offending article was written by Gordon Freeman, a 62-year-old chemistry professor at the University of Alberta, and appeared in June 1991 in a special issue of the journal devoted to the proceedings of the international conference on the Kinetics of Non-Homogeneous Processes in Non-Linear Dynamics. Entitled "Kinetics of non-homogeneous processes in human society: unethical behavior and societal chaos", the article consisted of Freeman's assertion that cheating had become more common among first- and second-year chemistry students, followed by his belief, derived from interviews with students, that they lacked traditional moral values because they had been neglected by their working mothers.

As editor of the conference issue, Freeman was able to place his article in the journal, but its publication also had the approval of the journal's regular editor, physicist Ralph Nichols of York University, Toronto. Bruce Dancik, editor-in-chief of NRC journals, saw it only after publication, and was outraged.

"It shouldn't have appeared. It wasn't science and didn't belong in a scientific journal", he said. "Anywhere along the way

it could have and should have been caught and wasn't. We published an apology about six months after the article appeared [but] after that we continued to get a lot of letters. Quite a few groups wanted us to reprint the journal without the offending article. That we refused to do, of course: we didn't want to rewrite history".

Critics deplored the fact there was no formal way for qualified social scientists to respond. Dancik's talks with the most persistent of them led to planning of a journal issue devoted to responses and rebuttals.

The monograph will appear next February in a format similar to the physics journal, and will contain articles by Pat Armstrong, chairman of the sociology department at York University (who calls the "so-called study" a joke); Margaret Eichler, professor of sociology in the Ontario Institute for Studies in Education; and Connie Stark-Adamec, president of the Canadian Psychological Association. It will also include a selection of letters the journal has received, along with some editorial responses. It will be sent to the 5,000 institutional subscribers in Canada and to the 1,100 physicists who received the conference journal containing Freeman's article.

Freeman, meanwhile, seems pleased with his notoriety. He began appearing on radio talk shows after his article was published, and newspaper advice columnist Ann Landers published his letter to her. "I've touched a nerve, haven't I?" Freeman told a reporter.

David Spurgeon

Squabbling over Australian science

Sydney. As in other recent elections in the English-speaking world, science and science policy are poised to play an inconsequential role in the Australian federal elections, which must be held within the next few months. With neither the ruling Labor party nor the opposing Liberal-National coalition in a strong position, the Federation of Australian Scientific and Technological Societies (FASTS), the country's main science lobbying group, had hoped to gain some political influence, but by attacking the science minister it has managed only to chill relations between the government and the scientific community.

A paper issued in August by the Australian Science Ministry and billed as a major science policy statement was recently described as "underwhelming" in a leading article in the FASTS journal. The article, written by executive director David Widdup, went on to declare that Ross Free, the science minister, was living up to his nickname

"agenda-free". Free responded by describing the FASTS board as "wimps", and repeating his earlier support for a "more representative" body to replace FASTS.

Ditta Bartels, the president of FASTS, has since apologized to Free, and acknowledged his efforts in persuading the federal government to retain a substantial research and development tax concession. But news of the retention of the tax concession had already been leaked to the media, and the only other identifiable initiative in the science ministry paper was for an additional policy emphasis on engineering.

A spokesman for Free said that the August policy statement was adequate and that no further utterances were planned. In the meantime, letters written to Free during the dispute indicated, according to the spokesman, that FASTS favoured the Liberal-National opposition and that the lobbying group had therefore lost all influence with the minister.

Mark Lawson

Hughes looks east for next expansion

Washington. Continuing its efforts to expand outside the United States, the wealthy Howard Hughes Medical Institute is weighing the idea of supporting researchers in the republics of the former Soviet Union and Eastern Europe. Purnell Choppin, the Hughes president, says that the institute is giving "very serious consideration" to establishing a five-year, \$14-million grants programme in nations of the ex-Soviet bloc.

The search for eligible researchers — a hard task in a research structure still dominated by the former Soviet academies and their bureaucracies — could begin early next year. It would mark Hughes' third expansion outside the United States in as many years. In 1991, Hughes issued grants to 24 researchers in Canada and Mexico, and this week it announced awards to another 29 scientists in the United Kingdom, Australia and New Zealand (see *Nature* 357, 101; 1992).

Choppin says the institute is at present discussing possible funding mechanisms with Russian scientists and with US agencies and science academies that have been working in the former Soviet Union. Until now, Hughes has resisted starting a grants programme in the area on the grounds that it would be too difficult to ensure that its rigorous scientific standards could be met.

But as an increasing number of Western organizations look towards Russian and Eastern European science, some techniques for picking the best researchers have started to emerge. Choppin says Hughes will be looking at such criteria as whether a scientist has published outside the Soviet Union, and in English — two factors that will simplify the review process, at least. The institute is also considering opening a small office in Russia to administer and monitor the programme.

Hurdles still remain, however. Choppin says that Hughes is still wrestling with such fundamental issues as Russian tax laws and the mechanics of transferring money into the region's rudimentary banking system. But if Hughes officials can work out the details during the next two months, they could get final approval for the new programme at the next Hughes trustees meeting in February.

Although the average grant in Hughes' international programme so far has been about \$90,000–\$100,000 a year per scientist, grants in the former Soviet Union and Eastern Europe would probably be smaller, but more numerous. According to some estimates, as little as \$2,000 a year will support a Russian researcher, although Choppin says that additional travel and equipment money is required for Russian researchers to keep up to Western standards. Even so, \$14 million can go a long way, and the new programme, if approved, could more than double the number of Hughes' international grants. **Christopher Anderson**

Universal Darwinism

SIR — In a recent book review (*Nature* 360, 25; 1992), Richard Dawkins comes close to admitting that evolution by Darwinian selection is so obviously *a priori* true in this and all other imagined worlds that it belongs to a category of knowledge that 'is more important than the facts that just happen to be true'; the latter being so numerous (blood is red, and so on) that no individual memory can cope with their retention. Whereas the more usual fun debate in biology is arguing the toss over whether life is an act of God or an act of natural selection, Dawkins is now dangerously close, in his quest for scientific respectability for his belief in Universal Darwinism, to bringing natural selection over to the side of God. It is no longer a scientific theory requiring all the usual rigours of proof and disproof but a set of beliefs on a par with theology which is *a priori* very likely to be true in all places at all times.

The problem with this view of life is that it makes a statement of probability ('very likely to be true') based on the unique singularity of life as we know it here on Earth. This is equivalent to making erroneous statistical calculations of how many times HIV should have transferred to the human species during the millions of years separating us from monkeys, given that a transfer has been observed to have happened uniquely in the past ten years.

If we define evolution as modification by descent, that is, a process for changing the genetic composition of a population with the passing of the generations, then there is no reason to suppose that, in other worlds at other times, processes that we actually know to be untrue here on Earth are untrue over there. Lamarckist evolution is a case in point. Mutationism is another. Both could be front runners in a mode of life with different procedures of inheritance and development. And there could be many more once we settle into our comfy armchairs and take to heart J. B. S. Haldane's view that "the universe is not only queerer than we suppose but queerer than we can suppose" (*Possible Worlds*, 1927). Even on Earth, there are other non-Darwinian processes contributing to the evolution of adaptive biological novelties which may (or may not) be operating elsewhere. If we stick to the only acceptable scientific attitude that true facts are those which are observed, then we will never know *scientifically* about imagined lifestyles elsewhere. All we can know in the meantime is via the hard experimental slog uncovering the ever-increasing number of facts of life as it is around us, including Earth-bound natural selection.

There is nothing rational or law-like about biological organization and the processes that gave rise to it, given that there are no predictable regularities of events on a par with physical phenomena. This should not be taken as a counsel of despair driving us towards the intellectual suicide of unknowable Universal Darwinism. On the contrary, the study of illogical and unpredictable organisms and the variety of bizarre and chaotic evolutionary processes at play should be an exhilarating challenge.

Gabriel A. Dover

Department of Genetics,
University of Leicester,
University Road, Leicester LE1 7RH, UK

Genuine genius

SIR — Jonathan Katz¹ complains about systems of competitive review which encourage "consensus science", and wants to reform the system so as to encourage "original and venturesome research". But I doubt that any system, conservative or reformed, can either silence or create genuine genius.

It is part of the nature of new and original ideas that present systems tend to discourage them. This was true at the time of Jeroboam in Israel, where King Solomon first tried to silence Jeroboam's ideas for social reform by offering him a lucrative job within the system². And it was true at the time of Descartes in Europe, where even the University of Leyden forbade mention of his name and where the atmosphere that led to Galileo's persecution led Descartes to decline to publish *Le Monde* during his lifetime³.

Katz suggests reforming the present funding system in science to encourage original and venturesome research, but the sort of people who are good at manipulating systems will probably quickly learn to succeed in whatever new system we adopt, while genuinely original and venturesome thinkers will probably always have a hard time. That is not necessarily a bad thing. The fact that an idea is original and venturesome does not make it true. The conservativeness of social and academic systems tends to weed out ideas that are original but not true, and helps to ensure that the original thinkers who survive the tendency of the system to discourage them will be those whose ideas are truest and most deserving of a wide and lasting hearing. The conservatives who tried to silence Descartes succeeded in silencing hundreds or thousands of lesser thinkers, but Descartes' ideas, 300 years later, are still influencing our physics, physiology,

geometry and philosophy.

Frank J. Leavitt

Jakobovits Center

for Jewish Medical Ethics,
Ben Gurion University of the Negev,
Beer Sheva, Israel

1. Katz, J. *Nature* 358, 10 (1992).
2. 1 Kings 11:40.
3. Russell, B. *A History of Western Philosophy* 559, Simon and Schuster, New York, 1945.
4. Hart, J. T. *The Lancet* i, 405-412 (1971).

More on legumes

SIR — Tate and Enneking's warning¹ of how consumers could be supplied with *Vicia sativa* in place of red lentils is not a unique case. A few years ago, yellow lentils were exported from Turkey to India and, because they resemble pigeon pea (*Cajanus cajan*), which is commonly consumed in India, they became popular. Traditionally, in the Middle East, yellow lentils are consumed after they are soaked in water and the washings discarded. But in India, the imported lentils were eaten without detoxification as the Indian population is accustomed to consume pigeon pea 'dal' without any processing. The import of yellow lentils from Turkey was therefore discontinued.

Another legume belonging to the same subfamily as *V. sativa* — *Lathyrus sativus*, which contains the neurotoxin beta-N-oxalyl aminoalanine (BOAA) — is cultivated in India and other countries². The neurotoxic effects of *L. sativus* have been well documented in several animal species and outbreaks of neurolathyrism in humans have been reported from India, Bangladesh, Nepal and Ethiopia. According to recent estimates, nearly one million tonnes of *L. sativus* are produced every year for human consumption. The government of Nepal recently banned its import.

In the quest to provide protein-rich legumes, particularly in regions where dietary protein deficiency is widespread, special care has to be taken to ensure that only legumes free from toxic constituents are encouraged. Because there is generally no restriction on the international trade of legumes, the Codex Alimentarius Commission of the FAO/WHO, as well as countries exporting, importing or cultivating legumes reported to contain toxic constituents, have to produce appropriate guidelines to safeguard human health.

Ramesh V. Bhat

T. C. Raghuram

National Institute of Nutrition,
Indian Council of Medical Research,
Hyderabad 500 007, India

1. Tate, M. E. & Enneking, D. *Nature* 359, 357-358 (1992).
2. Tulpule, P. G., Nagarajan, V. & Bhat, R. V. *Environmental Causes of Food Contamination* Environment India Series 1 (Department of Environment, New Delhi, 1982).

Reply to Menger and Haim

SIR — In the 10 years or so I have done editorial work for the *Journal of the American Chemical Society* (JACS), I have from time to time incurred the anger of authors whose papers I felt it my duty to reject, both with and without advice from referees. One colleague threatened a lawsuit, one accused me of forging a referee report, and I was once or twice treated to a selection of telephoned obscenities. The Commentary by F. M. Menger and A. Haim (*Nature* **359**, 666; 1992) represents the first case I know of in which the authors published a paper about not being able to publish their papers in JACS with sufficient ease.

Menger and Haim's article breaks new ground in other ways and introduces a new genre of scientific writing. This new genre can be called "post-modern scientific writing," as distinct from "naïve scientific writing", with which readers of other scientific journals and other authors are more familiar. In one example of this innovative mode, Menger and Haim present the following. "Schowen wrote: 'I don't think this paper is any good. Breslow and Huang never said they measured a negative rate'. It was difficult to understand the reply as Breslow and Huang had published an entire list of negative rate constants."

The naïve reader envisions Schowen aggressively rejecting a paper on a subject with which he is unfamiliar, perhaps leading to "considerable anguish" on the part of its author. Now I extend the Menger-Haim quotation from my letter by 14 words: "I don't think this paper is any good. Breslow and Huang never said they measured a negative rate, only that the rate decreased with imidazole concentration. They calculated a negative rate constant." With this new information, the author's anguish becomes itself "difficult to understand". Indeed, in the normative genre of scientific writing and reading, the attenuation of this quotation by Menger and Haim, combined with their interpretative sentence, might be taken as a "truly gross" misrepresentation. But this view fails to take account of Menger and Haim's post-modern style. Their quotation of my letter places the word *rate* directly before the quotation mark with no intervening punctuation. Under post-modern close analysis, the reader is alerted that the authors may have information that would reverse the reader's understanding if it should ever come to light.

With the post-modern scientific style, the reader must be deeply cautious about the surface text, and must probe analytically the grammar, syntax, punc-

tuation and every other aspect of form, always considering that other, deep readings may correspond to what the naïve reader would crudely describe as the 'truth'. Most important, much vital information may not be made available to the reader. For example, consider as a deep text that JACS rejected papers by Menger and Haim not merely because I was "defensive and evasive", but because the papers by Menger and Haim were frivolous, erroneous or so turgid as to cause more confusion than enlightenment. A naïve text might describe the laborious and public-spirited efforts of good-willed referees to help Menger and Haim formulate logical, honest arguments. These referees now find their confidential reports parodied by appallingly selective quotation in high post-modern style. A disadvantage of the naïve mode for Menger and Haim would have been a loss of self-serving ethical pretence, low drama, a sense of intrigue and entertainment value.

Naïve readers should note that Menger and Haim make much of my request that the *Journal of Organic Chemistry* (JOC) retract Menger's publication (56, 6251; 1991). I now see my action as arising from my own failure to appreciate the post-modern approach. I objected (making clear to the JOC editors that I was writing as a reader of JOC and not as an editor of JACS) to several aspects of the paper, but chiefly to Menger's Fig. 1 and the associated text. The caption of this figure reads (I naively cite its entirety): "Figure 1. The Anslyn-Breslow mechanism utilizing the five rate constants in eq 1 of their article² (and given in eq 1 of the present article). A steady-state treatment of intermediate I generates these equations." (The superscript '2' refers to E. Anslyn and R. Breslow, *J. Am. chem. Soc.* **111**, 4473; 1989.)

The mechanism in the figure violates the principle of microscopic reversibility, and Menger's paper attributes the violation to Anslyn and Breslow. Because the contents of the figure do not appear in Anslyn and Breslow and the verbal discussion in Anslyn and Breslow's paper explicitly excludes such a mechanism, naïve readers tend to see the figure and text as a misrepresentation of the content of the Anslyn-Breslow paper. This would be wrong: the close textual analysis required by the post-modern mode shows that the superscript '2' of the caption is not attached to the names "Anslyn-Breslow" nor to the word "mechanism" but rather to the word "article". Only a naïve reader will thus proceed incautiously and incorrectly to the conclusion that the material shown in

the figure is claimed actually to be present in the cited article.

Nature's launching of the post-modern genre of scientific writing is probably irreversible, so the attendant requirement for close analysis by scientific readers is of some concern. Chemists must already try to maintain currency in fields from materials science to molecular biology, and will now have to add to their obligations developments in the theory of literary criticism. Still worse will be the problem, for those journals that adhere to the naïve mode, of trying to ascertain the accuracy of submissions. Naïve 'accuracy' poses no problem for *Nature*, as the Menger-Haim publication illustrates. For the benefit of other journals, however, it might be hoped that *Nature*, which I believe has experience in the use of magicians to validate its papers, could make services in this category broadly available.

R. L. Schowen

Departments of Chemistry, Biochemistry
and Pharmaceutical Chemistry,
University of Kansas,
Lawrence,
Kansas 66045, USA

Germane reviews

SIR — Michael Ignatius's concerns¹ betray, at best, an excessively narrow view of the place of review articles in modern interdisciplinary research and, at worst, a fundamental misunderstanding of the methods employed in a machine literature search.

The salient question about his example — plasticity in the developing visual system — is not how many thousands of articles and reviews exist collectively concerning each individual subtopic of this topic, but how many reviews exist that are germane to the specific combination of subtopics in it. In using the NLM program *Grateful Med* to search Medline for the period 1986–92, I found 10,140 articles under "vision" of which 916 were reviews, and 1,479 articles under "neuronal plasticity" of which 320 were reviews. But only 7 review articles that contained both terms were retrieved^{2–8}, and even fewer were identified when more terms were included in the search parameters.

Manfred E. Wolff

Immunopharmaceutics Inc.
11011 Via Frontera,
San Diego, California 92127, USA

1. Ignatius, M. J. *Nature* **360**, 100 (1992).
2. Rauschecker, J. P. *Physiol. Rev.* **71**, 587–615 (1991).
3. Aoki, C. & Siekevitz, P. *Sci. Am.* **259**, 56–64 (1988).
4. Levine, S. M. *Am. Optom. Assoc.* **59**, 135–139 (1988).
5. Kasamatsu, T. *Prog. Brain Res.* **88**, 599–616 (1991).
6. Imamura, K. & Kasamatsu, T. *Jpn. J. Physiol.* **41**, 521–549 (1991).
7. Singer, W. J. *Exp. Biol.* **153**, 177–197 (1990).
8. Shatz, C. J. *J. Neurobiol.* **21**, 197–211 (1990).

IPCC strategies unfair to the South

Jyoti K. Parikh

The IPCC assessment of climate change is full of assumptions unfair to the developing countries. It is essential that revisions take these problems into account if effective strategies are to ensue.

BEFORE the Earth Summit, a framework for a climate change convention was formulated by an intergovernmental negotiating committee, which was eventually signed by many heads of states at Rio. Another parallel exercise was carried out by the intergovernmental panel on climate change (IPCC), which constituted three working groups whose reports are published in three volumes. Several hundred professionals were involved in preparing the three reports¹, the first two of which are intended to be objective scientific assessments. But the third, on response strategies, contains a lot of subjective assumptions, many of which do not do justice to the developing countries (the 'South'). This volume deals with how to reduce CO₂ emissions, in which regions and with what possible measures. Because CO₂ emissions are largely connected with the use of fossil fuels and to some extent deforestation and land degradation, the debate is about energy use in future by different world regions. Energy is an important ingredient for development and if fossil-fuel use, the main energy source of most developing countries, has to be restricted, development will be retarded or made more expensive. After examining which world region emits how much, IPCC projects probable CO₂ emissions models up to 2025 for different regions.

It is well known that of the 5.6 billion tonnes of carbon emitted during 1988, more than 70% is contributed by the developed countries (the 'North'). Because carbon accumulates over 100 years, the North's share in accumulated carbon since the industrial revolution is more than 85%; hence its responsibility in precipitating the 'greenhouse effect' is in the same proportion. The cost of CO₂ emission reduction in the United States alone is estimated to be in trillions of dollars for the next century. It is therefore essential that the South's need to develop is kept in view, together with the fact that the North has the major responsibility in bearing the financial burden of reducing CO₂ emissions.

IPCC world view

I discuss here the shortcomings of the IPCC results in the light of these facts. My comments are limited to the results up to the year 2025, the period relevant for policy purposes. Moreover, my discussion pertains only to fossil fuel-

related emissions and does not deal with deforestation-related emissions. I am limiting my comments to North America because Western Europe emissions are not high, and the turmoil in Eastern Europe makes predictions impossible.

IPCC uses a simple model to project future emissions up to 2025 based on past emissions as initial conditions and a set of future growth rates for different world regions. It calls a set of model results for different world regions a scenario. Obviously, the 'scenario' will depend on the inputs (future growth rates). Because no one knows the future growth rates, subjective assumptions are always involved. Table 1 is reproduced from the IPCC 'business as usual', or reference, scenario, which is a vision of the future in the event of no policy for global emissions. Fairness requires that the assumptions concerning future growth rates of different world regions for the reference scenario should be guided by considerations about whose responsibility it is to reduce emissions, who needs more energy for economic growth and development, and who has reached a stage where emission growth rates are already reduced.

Instead, IPCC assumes that the present inequalities among different world regions will increase considerably. Table 1 shows that annual per capita carbon emissions in North America will increase from 5.08 tonnes in 1985 to 7.12 tonnes in 2025, a 77% rise in total emissions and 40% rise in per capita emissions. This is inconsistent with falling carbon

emissions per unit GDP. The IPCC results show that in terms of per capita emissions, emissions in North America will equal 13.2 Africans or 11.1 South East Asians or 7.8 Latin Americans even in 2025. It is possible that emission growth rates will decline as a matter of course due to new technologies and conservation efforts. These expectations can be legitimately built into inputs for reference scenarios. But, as Table 2 shows, IPCC does this only for centrally planned Asia, and south and east Asia where emission growth rates decline compared to the past by nearly 30%. In North America, however, they increase substantially (by 1,200%!).

Emission cuts for all

Having incorporated this inequality and substantial increment in the inputs for reference scenarios, other runs are constructed entitled 'stabilization' or '20% reduction'. Stabilization is supposed to keep the annual rate of global emissions constant at the level of 1985 (5.15 billion tonnes). When emissions are added for all the regions, the global total reaches 12.43 billion tonnes in 2025 in the reference scenario. Because a substantial increase for North America is permitted, the required global emissions cuts are as much as 59% in 2025. It is obvious that the higher the consumption of the North, the higher the cuts for the South. But because considerable fat is permitted in the reference scenario itself, these cuts mean no sacrifice to the North. The IPCC document is unclear

TABLE 1 World view contained in IPCC reference scenarios

Regions	Total emissions (billion tonnes)		Per capita emissions (tonnes)	
	1985	2025	1985	2025
Global Totals	5.15	12.43	1.06	1.56
Developed (North)	3.83	6.95	3.12	4.65
North America	1.34	2.37	5.08	7.12
Developing (South)	1.33	5.48	0.36	0.84
Africa	0.17	0.80	0.29	0.54
Centrally planned Asia	0.54	1.80	0.47	1.15
Latin America	0.22	0.65	0.55	0.91
South and east Asia	0.27	1.55	0.19	0.64
Reductions required for				
Stabilization		(59%)		(32%)
20% Reduction		(67%)		(46%)

Figures are from Table 3.4 of ref. 1 (page 56). Only the relevant numbers are reproduced to save space. Developed countries (North) include North America, Western and Eastern Europe, OECD Pacific. Developing countries (South) include Africa, centrally planned Asia, Latin America, Middle East, south and east Asia.

TABLE 2 Comparison of IPCC growth rates with past trends

Regions	Past trends	IPCC annual growth rate	Difference in growth rates past and future (%)
Period	1979–1988	1985–2025	
North America	1.43	0.11	1,200
Centrally planned Asia	3.05	4.22	–29.7
South and east Asia	4.46	6.77	–34.1

Past trends are taken from Oak Ridge National Laboratory (ORNL) data and pertain to major countries dominating these regions, that is the United States, China and India, respectively.

IPCC growth rates are for the reference scenario, that is before cuts are applied. Stabilization would require further cuts.

about how these cuts are to be shared, but it mentions (p.69) that instead of a 3% growth rate, the emissions of developing countries could grow at 2% if they conserve energy. There are several places where it is suggested that *all (sic)* countries must reduce emissions, but the cuts for South are already built into reference models (see Table 2). Then there are additional cuts for stabilization due to increased emissions in North. If cuts are the same across the board, North America under the 'stabilization' model, despite 59% cuts, continues to emit almost the same per capita CO₂ in 2025 as it does at present! Thus, the stabilization scenarios of IPCC stabilize the lifestyles of the rich and adversely affect the development of the poor.

The stabilization scenario merely ensures the same rate of emissions as in 1985, that is 5.15 billion tonnes per year, and will still lead to additional emissions of 206 billion tonnes of carbon during the next 40 years. This is why IPCC also has a '20% reduction' scenario, requiring cuts in emissions by 67% in 2025. Once again, if the reference run had not been so generous to the North, cuts would not have to be so drastic for all.

It is illustrative to see how the South perceives the stabilization model. For example, when we worked it out in a limited exercise², we allowed India and China to grow at less than the 'business as usual' rates and found that the North would have to cut emissions by 30% in 2025 to offset increased emissions by only China and India so as to keep global totals constant as in 1986. Despite this growth, India's per capita carbon emissions would increase from 0.2 tonnes in 1986 to 0.6 tonnes in 2025, whereas the world average in 1986 was 1.1 tonnes.

The IPCC document shows great concern (to the point of paranoia) about 'potential growth' in emissions from the South, overlooking the fact that total emissions from the North, despite its smaller share of the global population, are higher than those from the South. The South needs time to develop, and the North must accept that, for a certain time, the burden of the stabilization scheme must be borne by all those and only those who emit more than the world average emissions in 1985, say 1.1 tonnes per person. This should be the

guiding principle of negotiations: if, however, the South finds it advantageous to reduce carbon emissions due to the benefits of reduced energy costs or pollution, it is up to it to do so. Such considerations must apply to the North and South alike. Of course, many developing countries will exceed this limit during the period and, if they do, they should also reduce consumption, provided the North has already done so.

So far, the discussions have been limited to annual emissions. But CO₂ remains in the atmosphere for more than 100 years, so responsibility for the concentrations accumulated due to emissions over at least this time needs to be worked out for all regions and countries. Such an exercise has been done³, but IPCC has not taken note of this very relevant suggestion.

Why do forecasts such as IPCC's get developed despite the South's participation in the debate? The answer is that the South is not yet ready for negotiations. First, if the South had been articulate enough to formulate its priorities for the Earth Summit, it would have placed on the agenda environmental problems such as unsafe drinking water, lack of sanitation facilities and unclean fuels which affect millions of people. But the South was persuaded instead to discuss the longer-term problem of climate change, for which it had not done adequate homework.

Post-Rio exercises of IPCC

All the IPCC figures need to be reworked to account for the various shortcomings I have discussed here. Post-Rio activities should include further analysis of the responsibility for climate change, who is affected and what is a fair mechanism to reduce emissions. Details of financial transfers need to be worked out: how much money is required, who should pay, who should receive and the criteria for payments and receipts. I have some further suggestions.

First, the reference scenario against which all future policies are compared should be designed carefully to avoid bias favouring the developed regions.

Second, stabilization scenarios should allow the developing countries to grow normally, at least till they reach world average emissions in per capita terms. The cuts should be imposed on only

those regions who are above world average in a base year (say 1990).

Third, throughout the exercise, accumulated emissions from 1950 (or some might say 1850) should be shown for each region. That should be the guiding principle even after 2025.

Finally, enforcement mechanisms as well as the financial transfer mechanism to the South should be kept at the centre of all models. Measures such as the annually tradeable emission quotas system, carbon tax, climate fund, off-sets, and so on need to be analysed in depth.

A tradeable emission quotas system may be superior for the following reasons. It would permit the North to emit more, provided it buys surplus (unused) quotas from the South and results in North–South transfers — not as aid or charity but as trade. It would provide economic incentives to the South to emit less CO₂ from the beginning, because the South could sell its saved quotas. The system will provide a greater incentive for developing new technologies and will also provide a market-determined way to value such technologies. Grace periods will keep the South backward, as it will have no incentive to adopt cleaner technologies and will not think of ingenious ways to reduce emissions and stay ahead at the CO₂ saving game. Yet in the absence of tradeable permits or other financial mechanisms, this may be the second-best solution.

At present, instead of a fair and market-oriented tradeable quota system, a carbon tax appears to be the favoured measure of the OECD countries to reduce fossil fuel consumption. But a carbon tax in the North alone would be pointless because carbon-intensive activities will shift to the South and global emissions will not be reduced by much. A carbon tax would make sense only if North–South financial transfers were an integral part of the strategy.

It is to be hoped that the post-Rio activities of IPCC will be constructive and fair to the South. In the end, it is quite likely that those involved in constructing the original models were not conscious of the inequality and injustice embedded in them. But now there is no excuse for failing to formulate fair policies for the future. □

Jyoti K. Parikh is at the Indira Gandhi Institute of Development Research, Goregaon (East), Bombay 400 065, India.

1. IPCC Climate Change: The IPCC Response Strategies (WMO and UNEP, 1991).
2. Parikh, J. et al. Consumption Patterns: The Driving Force of Environmental Stress (Report prepared for United Nations Conference on Environment and Development, IGIDR, Bombay, 1991).
3. Smith, K. *AMBIO* 20 (1991).

ACKNOWLEDGEMENTS. I thank K. Parikh and M. Grubb for comments. Supported in part by Climate Action Network.

Heading into an uncertain future

Economic difficulties, and the challenge of entering the European Communities, are each prompting Europe's northernmost countries to reassess their approach to the organization and funding of science

THE Nordic countries are not an alternative Europe in miniature, as first appearances may suggest, but are intrinsically a part of Europe. Occasional visitors to those parts must repeatedly remind themselves of that simple truth. But why should there still be confusion on this score?

The most obvious (and most trite) explanation is that the northern cities have a distinctive feel which is different from that of central and southern Europe's generally busier and less polite cities. The Nordic countries are also made distinctive by their physical scale; rotate Norway around its southernmost tip, and cities previously north of the Arctic Circle end up half-way between Milan and Rome.

At the same time, in terms of population, the Nordic communities are small. Sweden, the most populous, has fewer than 9 million people. Even with Iceland (not covered in this survey), the Faroes and Greenland, there are no more than 25 million Nordic people altogether.

The more important cultural difference between the Nordic states and the rest of Europe is (or used to be) political. For 40 years, Denmark, Norway and Sweden have been unique social democracies built around the principle that everybody who wants a job should have one, or alternatively the wages he or she would earn. The high costs of the world's most generous welfare programmes were sustained, until the mid-1980s, by high taxation and relatively rapid economic growth.

Now that has changed. Economic growth has faltered and high tax rates have become unpalatable, while the numbers of people without jobs have grown to make welfare payments an embarrassment. Scandinavia's alternative

route to socialism seems, at least for now, to have failed.

Finland has recently fallen into a similar economic mess for different reasons. Since the early 1950s, Finland was tacitly accepted by other Nordic governments and by the West to be part of a neutral buffer zone between what was then the Soviet Union and the outside world. Much of Finland's prosperity in this period derived from that unique trading position. But the successors of the Soviet Union do not need the services Finland used to provide, and cannot afford to pay for them in any case.

This sudden and unexpected change of economic fortunes is a large part of the reason why Nordic governments have been looking eagerly southwards, to an arrangement with, and eventually membership of, the European Communities (EC). Despite last Sunday's referendum in Switzerland, the Nordic states (and Austria) are still planning to become part of the European Economic Area, a general free-trade area centred on the EC.

The same enthusiasm for the wider Europe has marked the planning of Nordic research for the past several years. All governments have sought to participate as fully as they can in collaborative European projects, not just EC programmes but more widely based projects such as those of EUREKA and of the European Science Foundation. Unsurprisingly, the Nordic states have been widely welcomed as partners in research. In the discussions with working researchers on which the following survey is based, people have frequently described their pleasure that the old patterns of transatlantic collaboration are being replaced by collaboration within Europe.

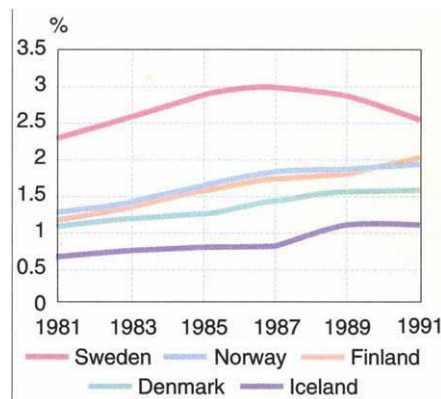
The research communities of these Nordic states are cheerful in other ways. Most scientists, for example, say that there is no shortage of funds for travelling to meetings elsewhere, either for themselves or their students. Generally, independent investigators are also content with the scale of support for basic research, at universities and elsewhere. Indeed, the Nordic research community gives the impression of being uniquely contented with its lot.

But, as this survey also demonstrates, there are changes in the air. The research communities of all four Nordic countries (though to a lesser extent in Norway than in the other three) are now facing two types of external pressure to which they

are not accustomed: economic constraints which have, for the time being, curbed any thoughts of further expansion in research budgets (meaning that new projects can only be started if old ones are abandoned). And a political emphasis on the primary role of the market in general, and the European market in particular, as a customer for the "products" of research.

The broad result of these twin pressures (familiar to anyone living in Britain's in the 1980s) has been a shift in official thinking from a "welfare state" to a "market state" ideology. Under these pressures, each of the Nordic countries is, to a greater or lesser extent, currently transforming the way that it supports research and development.

The main goal of this restructuring — at least as it is expressed by science administrators — is to make the research base leaner, fitter and more responsive to the needs of industry. One way of doing



Research expenditure has been climbing steadily as a proportion of GNP in all Nordic states — apart from Sweden — over the past decade

this is by encouraging more competition for funds. Another, at least in principle, is by developing new institutions, such as university-based centres of excellence, that will make it easier for government-funded research to be guided by the needs of, and on completion rapidly transferred to, the private sector.

How effective this restructuring will be remains to be seen. For many scientists express concern that, in their enthusiasm to cut all forms of public spending — as well as remain true to an ideological commitment to the market — their governments may be tempted to remove from research the protection which they

Any visitor to a region as geographically scattered and culturally diverse as the Nordic countries must unfortunately be selective. This supplement, written by John Maddox and David Dickson, does not pretend to do full justice to the wide range of research to be found in the Nordic countries, much of an extremely high standard. Rather than attempting a comprehensive survey, we have tried to highlight some of the major policy issues which are currently pre-occupying the scientific communities in each, and to illustrate these issues with selective descriptions of individual institutions and research groups. We thank all those who generously gave up their time to help us in this project — and apologize to those who, for reasons of space, we have not been able to mention in detail.

have given it in the recent past. Yet such protection may remain necessary if fragile economies are to prosper without the "demand-pull" that the welfare state has previously provided.

Take, for example, the question of government support for technological research (particularly in universities). In the 1980s, this became a major field of expansion, as all Nordic governments introduced national technology programmes designed to ensure the build-up of expertise in areas (such as biotechnology, microelectronics and materials science) considered essential to the future health of their industries.

The results of these programmes were mixed. In some cases, the quality of output has been high. In others, often due to a lack of adequate assessment of original grant proposals, it has been disappointing. And even where the research results have been impressive, they may (for reasons of conservatism or indifference) not have been picked up by the industry which they were meant to benefit.

Unsurprisingly, there has now been a backlash against the idea of researcher-led technology programmes. Industrialists are being put in the driving seat, with firm instructions to ensure that these programmes achieve an appropriate "output" in the form of products that can compete in the international marketplace. A well-intended goal, perhaps. But although industrial research managers may be better at identifying

their short-term research needs, there is no guarantee that they (or their directors) will be any better than academic scientists in foreseeing long-term technological opportunities.

Enthusiasm for Europe is also tinged with apprehension. But faint hearts are not likely to win the day. Each country now needs a bold strategy for its research community. One goal must be to find ways to build on pre-existing strengths, for example on the long tradition of medical research in Sweden (the basis of its strong pharmaceutical industry) or the more recent development of oil research in Norway (which has provided an across-the-board boost to the country's research efforts).

A second need is for each country to tap the traditional inventiveness of its populations to carve out niches in the global economy. For some, this will mean finding radical new approaches to products that can be generated from natural resources (such as fish in Norway, or paper and pulp in Finland). For others, it could mean taking adventurous steps in new directions (as the furniture design and mobile telecommunications indus-



tries have done with such success).

The Nordic populations have a long tradition of thriving under harsh conditions. This in itself may provide their scientific communities with the skills required to prosper in an increasingly competitive environment. If they can keep their politeness and good-humour intact in the process, the rest of us will be the richer as well. □

Collaboration: yes, but in which direction?

WILL entry into Europe mean the end of efforts to increase inter-Nordic cooperation in research? In the short term, most people agree that the impact will, indeed, be negative. Looking longer ahead — or perhaps just being more optimistic — the argument is that, once membership of the European Communities (EC) is established, regional cooperation will become as important as ever.

Close cooperation between the Nordic countries has long been a goal of politicians. Science has benefited in a number of ways. In particular, funds from the Nordic Council of Ministers have been used to establish various joint research institutes, perhaps the best known being the Nordic Institute for Theoretical Atomic Physics (NORDITA).

Today, with all eyes on Brussels and the terms of membership of the EC, pressure for closer Nordic collaboration has taken a back seat. "There is no real policy discussion going on about the future of Nordic cooperation," says one Swedish official. "It is just not on the agenda at present; all attention is focused towards Europe."

One body feeling the squeeze is the Nordic Industrial Fund. This was set up in 1973 by the governments of Denmark, Finland, Iceland, Norway and Sweden to boost technological and industrial development, primarily by identifying and financing Nordic research and development projects.

In the mid-1980s, NIF launched a number of major technology programmes on topics such as biotechnology, materials technology and information technology, in each case put together along the lines of similar programmes being run by the European Commission from Brussels.

Per Gjelsvik, the director of NIF, admits that the fund will now have to change course. "We have reached the conclusion that we will no longer be able to support the same research as the EC," he says. "So we will support research which will provide direct benefit to the Nordic countries, for example, in meeting the needs of their paper or fishing industries."

Gjelsvik is sure that, once Norway, Finland and Sweden become firmly em-

bedded in the EC, there will be renewed calls for regional cooperation. "Everyone is now talking about Brussels. I am saying, let's wait for a few years; when they see the importance of Nordic cooperation, they will turn back again."

In contrast to applied research, collaboration in basic science is continuing to grow, primarily because this is considered to be a cultural activity. In particular, the Council of Nordic Ministers has recently established a Nordic Academy for Advanced Study (NORDAS) to support postgraduate training and the operation of scientific networks.

"There are many small universities in the Nordic countries which have difficulty in setting up strong research groups and providing postgraduate training," says Leif Westgaard, executive director of the academy who works from the offices of the Norwegian Council for Science and the Humanities (NAVF) in Oslo. "Our idea is to pool these resources together to provide joint doctoral training; this year, for example, there will be about 40 courses, in subjects ranging from classical archaeology to planetary physics." □

Sweden: The times are a-changing

ONCE every three years, the Swedish government publishes a bill setting out its future priorities for science. The last one, produced at the beginning of 1990, anticipated a period of rising research budgets and expressed a desire for "continuity and renewal" in relation to the policies of the previous decade.

The next one, due to be published in February, promises to be radically different. For one thing, the Swedish economy has since plunged into crisis. For another, the government has changed political complexion, and the Social Democrat ideology which permeated thinking about science and technology during most of the 1980s, and indeed the post-war period in general, has given way to a neo-liberalism that emphasizes the primacy of the market and the desirability of cutting public expenditure.

At present, Swedish scientists have little to complain about. Sweden still spends almost 3 per cent of its gross national product (GNP) on research and development — a figure that puts it with Japan, Germany and the United States at the top of the list of countries belonging to the OECD. Despite the country's economic difficulties, basic research has so far escaped virtually unscathed, and this year's budget for the research councils is likely to remain at roughly the same level as last year's. Furthermore, Sweden still dominates Nordic science, both in terms of the money it spends and the quality of the science it produces.

But some cracks in the edifice are beginning to show. At the end of the 1980s, after a period in which it had been growing at more than 10 per cent a year, research spending by the private sector began to fall. Although government funding for research continued to increase, the result was still an unprecedented drop in the country's total R&D spending from 3.04 per cent of the GNP in 1987 to an estimated 2.54 per cent last year.

The government claims that the complex structures established to support research and development in the past — in particular the idea that applied research should be heavily supported by public funds and carried out primarily in universities — are no longer appropriate to the new competitive environment of the 1990s.

It is now arguing the case for centres of excellence in universities and for extra funding to be given to university research groups prepared to orient their activities towards the interests of Swedish industry; in short, for more explicit selectivity and elitism.

"We are saying that competition and

elitism are good for science", says Håkon Eriksson, professor of reproductive endocrinology at the Karolinska Institute in Stockholm, deputy secretary of the Medical Research Council, and secretary of a committee set up by the government to prepare a new science policy. "In a country like Sweden, to come out and say we support elitism and should put special resources into centres of excellence is something completely new; even if people had thought it, they had not dared to say it."

The modern system of R&D funding has grown hand-in-hand with the expansion of the country's welfare state. In particular, the rapid expansion in the 1950s and 1960s of the Swedish infrastructure, including power supplies, telecommunications and health services, provided both industry and the academic community with an important source of research funds (and formed the basis on which Swedish industry was able to make its substantial achievements in high technology).

It is this "Swedish model" of research and development that is being actively rejected by the new government. "There is now a contraction in the public sector, and heavy investment by the government is therefore not as important as it has been for Swedish industry," says Bjarne Kirsebom, under-secretary of state in the Ministry of Education and Science.

Exploring new ways of boosting industrial research without involving heavy public subsidies has been one of the priorities facing the new Minister for Education and Science, Per Unckel.

One goal that the government has already adopted is to double the number of PhDs produced in technical subjects by the end of the century. The hope here is that when these postgraduates move into industry, they will become the research "purchasers" of tomorrow.

Other goals will be presented in the bill that the government will submit to parliament in February, setting out the main guidelines for its proposed science policy over the three years 1993–96. The details are still being vigorously debated. But some preliminary ideas of the directions in which it is likely to point have already been published by the government in a recent discussion document.

Not surprisingly, the document has a strong neo-liberal flavour. It summarizes the overriding aim of the government's proposed strategy as being "to contribute towards the creation of long-term opportunities for a competitive Swedish knowledge environment", and takes a dig at previous government policies by stating a "general rule" that "research of

outstanding quality cannot be planned into existence by direct political means."

Some of its detailed proposals have already been controversial. For example, the document suggests that too much money may be going into agricultural research, and that a transfer of funds to other research fields might be "amply justified".

More widely debated has been the strong emphasis being placed on the need to build links with industry. "Universities have a key role as independent generators of knowledge, but we have to look on that knowledge as a key resource for society, and we have to improve that transfer of that knowledge to society", says Eriksson.

For some, however, such remarks suggest that the government is in danger of neglecting the specific needs of basic research by emphasizing what one scientist describes as "applied science for today's industry, not the basic research for tomorrow's".

Others are more concerned that, despite all its enthusiasm for closer links between industry and universities, the government has so far given little concrete indication of how it intends in practice to achieve this while reducing the use of public funds.

Much uncertainty, for example, hangs over the fate of the National Board for Industrial and Technical Development (NUTEK), the successor to the National Technology Board (STU). In the 1980s, STU played a key role in supporting technological research in universities in areas considered important to the country's future.

NUTEK is asking for a considerable increase in funding next year, much of which would be used to set up university-based "centres of competence" in areas of strategic industrial interest. The government has indicated that it is sympathetic to the content of NUTEK's request. But it still appears uncertain about how far its principles will allow it to go in committing public funds to support the interests of private industry.

The government's final decision on this point, as on others "still under negotiation", will not become known until budget figures are published in the science policy bill. "This is going to be probably the most interesting science bill we have had for many years," says Rikard Stankiewicz of the Research Policy Institute at the University of Lund. "There will be real change. But exactly what they will be — and whether they will be for good or ill — we will not know until we have seen the figures." □

Business-like universities

SWEDEN'S universities have become swept up in the government's enthusiasm for privatization. Later this month the Parliament is due to give final approval to a new higher education law, coming into effect next year, under which virtually all the internal administration of the universities will be freed of state control.

The key words in the government's justification for this radical move are expressed in the title of the original bill: *Quality through freedom*. Decentralized responsibility, it hopes, will encourage greater competition both between and within universities. And both will give universities and their faculty members the incentive to raise the quality and social relevance of their research. "We are changing from a collective to an individual system" as one university scientist puts it.

Not everyone is enthusiastic about the change. Some researchers are concerned that the result will be increased vulnerability to short-term pressures, in particular from Swedish industry. Others fear that the concentration of effort on specific research areas through the creation of centres of excellence (one of the government's key objectives) will reduce the scope for cross-disciplinary cooperation. But the government is determined to establish the law as part of its strategy to shift from a "welfare state" to a "market state" model for the economy.

There are six main universities in Sweden: Göteborg, Linköping, Lund, Stockholm, Umeå and Uppsala. At present, all are run according to procedures established by the state in the 1960s and 1970s. For example, they are required to conform to legally determined forms of internal organization, teaching salaries are negotiated centrally with the state, and parliament has to approve the creation of new university chairs.

In future, such decisions will be made locally. The government will appoint the board of a university (including industrialists) and the university's president and rector. But these will in turn be responsible for the way in which the university is organized.

There will be also be a big change in the way in which public resources are allocated to universities. Previously, this was done centrally by a government body, the National Board of Universities and Colleges (UHA). The board, however, has now been abolished. In future, each university will negotiate its annual grant directly with the government, based on a contract under which the university will promise to provide a certain number of undergraduate and courses in specified subjects.

It will then be up to the individual

university itself to decide how to meet the terms of its contract in the most cost-effective way. A university's research activities will be agreed in a similar way. But rather than detailed targets, the contract with the government will specify objectives, for example that the university concerned will perform research in particular fields, produce a specific number of postdoctoral students over a certain period of time, and so on.

Government officials claim that the changes will make it easier for universities to work in what they describe as a "business-like" way, and that this in itself will make it easier to respond to the needs of Swedish industry.

They also hope that, by emphasizing research excellence and productivity in the way that contracts are awarded, they will encourage universities to use the same approach internally. "For example, we are trying to find ways of stimulating quality in science by giving special bonuses to those who turn out more high quality PhD students," says Håkon Eriksson of the Karolinska Institute, who has been advising the government on its new research strategy. "The signal we are giving is that universities should develop similar ways of working."

The need for reform is generally recognized in the universities. There is a widespread feeling that many university departments have grown unwieldy, and that their administration needs to be streamlined, for example through organization into larger administrative units and thus fewer coordinating committees. The new law will make such changes easier to implement.

But there are also concerns about some of the implications of the government's policy. "It is a good thing that the government wants to encourage excellence in research," says Sven Kullander, professor of high energy physics at the University of Uppsala, and dean of the faculty of mathematics and physical science. "But we don't like the idea of centres of excellence, since there is a danger that these will become cut off from the cross-fertilization that is so fruitful in broad-based university departments."

Others claim that there seems to be a contradiction between statements by the government that it wants universities to have more autonomy, and its desire that their activities become more directly linked to the needs of industry.

On the political front, the reforms have been opposed by the Social Democrats in parliament, who argue that they will encourage elitist attitudes and undermine Sweden's traditions of egalitarianism. The government, however, is making no apologies about doing precisely this. □

Support from the workers

IT'S not often that a pot of gold falls into the research community's lap. But this could happen in Sweden early in the new year, following a promise by the government to distribute up to one third of a SKr20 billion (£2 billion) fund, built up in the 1980s through a levy on industry, to support industrially related research.

The so-called Workers Fund was set up in the early 1980s by the Social Democrat government of the time as a tax on salaries. Since the money raised was subsequently invested in the stock market, many saw the fund developing into a political lever over Swedish industry. So when the present conservative-led coalition came to power in October 1991, one of its first moves was to fulfil a pre-election pledge to drop the requirement on companies to contribute to the



New science minister Per Unckel — seeking new ways of paying for research.

fund, and to distribute its proceeds.

Precise details of how the research community will benefit are due to be published in February in the three-year science policy bill. Government officials say their preferred solution is to use the money to endow four or five private research foundations, each of which would be responsible for distributing the investment income (and perhaps some of the capital itself) to support projects in a particular field of research in ways that are of interest to Swedish industry.

Since the original promise was made earlier this year, the government has received a steady stream of proposals about how the money should be spent. The Swedish drug industry, for example, has submitted a list of ideas for long-term research projects, ranging from drug delivery systems to glycobiology. The Medical Research Council has also submitted a list of ideas (and has subsequently come under fire for the extent to which some of these appeared to reflect the interests of individual council members).

Overlapping with many of these proposals, the National Board for Industrial and Technical Development (NUTEK)

has proposed using the money to set up 30 university-based centres of excellence in fields of interest to industry. Like the proposed Faraday centres in the United Kingdom, these would carry out both postgraduate training and research, and act as a point of contact between universities and industry.

Per Unckel, the Minister of Science and Education, has been an enthusiastic supporter of the use of the Workers Fund to support such activities, particularly

because of the flexibility that private foundations would have, and is defending the proposal vigorously against cabinet colleagues who want to use the money for other purposes, such as pensions. But the dismantling of the fund is being actively opposed by the Social Democrats in parliament. And its value has been falling precipitously in step with the declining stock market. So although appetites in universities have been whetted, no one is taking the money for granted □

Sweden's central role

THE geographical isolation of Nordic scientists has long provided an incentive for collaboration with colleagues in other countries, both within Scandinavia and further abroad. A recent study carried out for the Nordic Council of Ministers has shown that the volume of inter-Nordic collaboration continues to grow with Sweden as its central focus.

"However, this growth has been less rapid than that of collaboration with scientists in the United States and in the countries of continental Europe," says Gunnar Sivertsen of the Institute for Studies in Research and Higher Education in Oslo.

The analysis was based on scientific articles published in the period 1975–90 and listed in the *Science Citation Index*. It was carried out by Sivertsen together with Terttu Luukkonen of the Academy of Finland in Helsinki and Olle Persson of the University of Umeå in Sweden.

For each of the four Nordic countries, the proportion of articles co-authored with scientists in other Nordic countries increased steadily during the period in question; in the case of Finnish scientists, the proportion of such papers almost doubled.

Collaboration has been particularly strong in clinical medicine and biomedicine, rather than the "hard" natural science of physics, chemistry and mathematics. And the importance of Nordic collaboration varied. Norway has the largest proportion of articles co-authored with at least one other Nordic country, and Sweden the least.

Detailed analysis of the bilateral links between Nordic countries revealed the extent to which Sweden continues to play the central role in Nordic science. Swedish scientists published almost as many papers during the period studied as those from the other three countries put together. And scientists from Denmark,

Norway and especially Finland cooperate with colleagues in Sweden — particularly in the medical sciences — more frequently than would be expected from the size of the respective scientific communities alone.

Yet, important as inter-Nordic cooperation remains, the bibliometric analysis shows that it is being overtaken by cooperation with non-Nordic scientists. During the period 1975–90, there was a relative decline both in cooperation within the Nordic region and with Great Britain. In contrast, both the United States and the countries of continental Europe have become increasingly important as collaborative partners.

Another bibliometric study carried out by Sivertsen underlines the role that the publication of journals has played in the integrating Nordic research into international science. The first Nordic scientific journal, *Acta Medica Scandinavica*, appeared in 1869. Initially published in German, the scientific language of the time, it now appears in English under the title *Journal of Internal Medicine*, although it still has a Scandinavian chief editor.

Today, there are more than a hundred journals that can claim to be Nordic in origin — that is, are edited in one of the five Nordic countries, but the majority of whose articles are written by scientists outside the country of the (chief) editor. "These journals are part of the process of the internationalization of Scandinavian science," says Sivertsen.

"The standard of articles is high; they are cited with the same relative frequencies as other scientific journals. And something which is usually seen as a drawback — the fact that we do not belong to a large language group — has in this case stimulated the use of English language publishing."

Many of the journals are rapidly becoming more international in their content. "The editors find that the best articles by Scandinavian scientists are no longer submitted to the journals if they do not do it. Going through nine years, we can see that the proportion of articles published by Scandinavian scientists in

Intellectual inward investment

EVER since the Middle Ages, Sweden has welcomed foreign scholars to its universities. The benefits have been mutual, since the Swedish academic community has used these visits to keep in touch with the outside world of scholarship. But many of today's scientists find it difficult to work separated from their research teams, and are reluctant to stay in a foreign university for any length of time on their own. So, rather than just inviting individuals to visit, why not complete research groups?

The idea has already caught the fancy of the Swedish government, first stimulated by proposals from the government of Japan to set up a branch of the Karolinska Institute in Tokyo, and subsequently reinforced by a recent visit by the Minister of Education and Science, Per Unckel, to Stanford University in California.

"Why not establish a permanent research facility of Britain's Medical Research Council on the campus at Uppsala, or one from Stanford at the Karolinska Institute in Stockholm?" asks Bjarne Kirsebom, undersecretary of state in the Ministry of Education and Science. He adds: "Ideas in this area are very likely to get government support."

The way the government sees it, implanting a high-level foreign research team on a Swedish campus would be an effective way of reinforcing the scientific strength of domestic research groups, for example through constant contact with scientists working in these groups, and the exchange of postgraduate students. One idea is that government funds would be used to create a research laboratory on a Swedish university campus which would be run by a foreign research council or other organization, and occupied by rotating groups of foreign scientists.

The proposal is still in embryonic form. But the idea of providing permanent bases for foreign research teams is already generating strong support in both the academic and industrial research communities. Kirsebom says that providing money for such facilities is precisely the type of use he has in mind for the proceeds of the Workers Fund (see opposite), and is hoping that his ideas will be included in the white paper on science policy, due next year. □

Nordic Collaboration in Science — a Bibliometric Study. Terttu Luukkonen, Olle Persson and Gunnar Sivertsen. Nordic Council of Ministers, 1991. Nord 1991:28

Internationalization via Journals. Gunnar Sivertsen. Nordic Council of Ministers.

these journals has fallen from 60 to 42 per cent," says Sivertsen, reflecting the fact that Nordic scientists are publishing an increasing number of their articles outside Scandinavia. □

Change at the Karolinska

ALONG with the Mayo Clinic, the Karolinska Institute in Stockholm is arguably the world's best known freestanding medical school. Founded early in the nineteenth century when defects in the Swedish Army's arrangements for treating wounded soldiers became apparent, the school now takes in 280 would-be physicians every year, together with more than a hundred potential orthodontists and a larger number of physiotherapists. Yet, as things have turned out, research and graduate training have become dominant, accounting for 80 per cent of the institute's budget. In total, there are 1,500 research students, not all of whom are physicians as well, and a similar number of full-time scientific staff.

The concept of a freestanding medical school is unusual. At Aarhus, Denmark, the university is the product of the recognition that a medical school cannot easily stand alone. But Stockholm is different, with two well-established universities and thus no shortage of expertise for teaching nonmedical subjects to would-be doctors.

The institute proper has nevertheless grown over the decades haphazardly. Only in the past few years has the administration taken a rationalizing axe to the proliferation of departments — more than a hundred of them — reducing the number now to fewer than three dozen. The principle had been that an able person recruited to the institute would be established in his own department, often physically separate from existing buildings. A consequence of the rationalization now under way will be the need to house larger groups of people in larger buildings, whence the seemingly endless construction work visible on the site of the northern campus 8 km from central Stockholm.

As part of the reorganization it is also agreed that most scientists should now compete for research funds with other applicants to the research councils from universities and elsewhere. The system has not been in place long enough for there to be a groundswell of discontent from disappointed applicants frozen out of research prematurely in their opinion. Despite the growing strength of the life sciences at other institutions in Sweden, it is likely that the Karolinska will be able to get more than its arithmetical share of the funds available.

Prestige helps. At the beginning of the century, the Karolinska Institute accepted what many thought to be the poisoned chalice of making annual nominations to the Nobel Foundation of prize-winners in physiology and medicine. In the event, the duty has proved more interesting than onerous. Indeed,

because the institute is equipped by the Nobel Foundation to keep up with current developments in physiology and medicine throughout the world, its function as a listening post for Nobel prizes seems also to be a useful adjunct to the administrative task of plotting a strategy — at the very least, it keeps senior people's ears close to the ground.

The Nobel connection also lends prestige, which is one reason why the northern Karolinska campus has an uneasy relationship with the new campus at Huddinge, on the other side of Stockholm, where there is a brand-new hospital and perhaps more striking, a Centre for Biotechnology.

Biotechnology has no obvious place in a medical school as distinct from, say, a pharmaceutical manufacturing plant. The biotechnology centre at Huddinge owes its existence to Professor Jan-Ake Gustafsson and to the Stockholm City Council, which has provided support for the centre since its foundation in 1985.

In its way, the centre is meant as a piece of Swedish social engineering. The southern suburbs of Stockholm are traditionally relatively poor — they are where newly arrived immigrants tend to live, for example. Part of the city council's objective has been to improve job prospects for these people and to enliven a part of the city previously empty of work spaces.

The result, so far not complete, is

modestly spectacular. The Centre for Biotechnology now lives in what is called the Novum Building, a glass building with strips of pastel-coloured plastic marking joints between services and with preparation rooms as spotless as operating theatres. The staff, while no longer growing quickly, works at a series of fundamental problems relevant to more applied ventures in biotechnology. One project, for example, is to use the Semliki Forest virus as a means of generating proteins of predetermined structure as well as a vector for transfecting naive cells with unspecified genes.

On the face of things it seems unlikely that the Karolinska's reputation will in any way be tarnished. Whether the basic research will spawn the sheaf of biotechnology companies that people originally hoped for is a still untested proposition, even though the city council has set aside a large tract of land still further south than the Novum building.

Gustafsson says that he nevertheless finds the interaction with the city council both necessary and salutary. It is necessary to keep the council happy with the research programme under way but at the same time it is valuable to learn from elected politicians of the enthusiasm their taxpayers hold for what remains a somewhat daring venture. In all the circumstances, the Centre for Biotechnology seems set for at least as many years ahead as the seven it has enjoyed so far. But prudently, its Novum building is apparently leasehold, not freehold. □

Following ancestral footsteps

MICHAEL Sohlman, the new director of the Nobel Foundation, appointed last May, has hired a steel band from Copenhagen to play at the banquet after this month's Nobel prize ceremonies. The band is in honour of Derek Walcott, the Caribbean poet who is this year's winner of the literature award. Sohlman plans no other innovations this year.

Sohlman's appointment to the Nobel Foundation is nevertheless of great interest. His grandfather worked for Alfred Nobel as an engineer in Paris at the turn of the century. On Nobel's death, he took charge of the operation of collecting together Nobel's assets in France and of transporting them to Sweden. The operation entailed the purchase of a Fiat motor car and two pistols to keep potential robbers at bay. The younger Sohlman says that this hazardous operation was nevertheless entirely legal. By his account, French law at the time specified that a person's domicile was the place at which he kept his horses which, in Nobel's case, was Sweden, not Paris.

The younger Sohlman, who describes himself as a "simple Treasury man" says that he sees no need for radical changes at the foundation, which is in any case



Nobel director Sohlman: sees no need for radical changes.

tightly circumscribed by Nobel's will. But he regards the relationship between the foundation and the nominating organizations specified in Nobel's will as crucial to the success of the operation, noting that the foundation itself is not able to make accurate scientific judgments but that it can monitor the way in which the nominating bodies go about their task. □

Merging for strength

MANY success stories have modest beginnings. Back in the mid-1970s — or so they like to tell it in Stockholm — a small Swedish venture capital company called Kabigen agreed to buy the air-ticket of a young Californian entrepreneur invited to Sweden to discuss collaboration on a joint research project.

The entrepreneur was Robert Swanson, then in the process of getting Genentech off the ground with Herbert Boyer of the University of California, San Francisco. Swanson had been invited to Sweden by the company's research director, Bertil Åberg, to discuss ways of applying genetic engineering techniques to a product for which the parent company, Kabi, had long been known: human growth hormone.

The research contract was one of the very first signed by Genentech. It resulted in the biosynthetic hormone Genotropin, which has since become a highly successful product; last year, Genotropin claimed 60 per cent of the world market outside the United States, with global sales of SKr2.00 billion (£200 million).

The drug has, in the process, become the best seller for a new company, Kabi Pharmacia, formed in June 1990 when Kabi merged with Pharmacia to create what is now one of the largest pharmaceutical companies in Europe (and is itself a subsidiary of a giant conglomerate, Procordia, put together under the chairmanship of Pehr Gyllenhammar, president of Volvo).

Like Kabi, Pharmacia has also had a long tradition of turning basic research carried out by university scientists into successful products. In the late 1960s, for example, the company, based in Uppsala, combined two lines of research at the nearby university — the discovery of immunoglobulin E and the development of radioimmunoassay technology — to develop a new technique for allergy testing.

The allergy test went on sale in 1974. Seventeen years later, Kabi Pharmacia has a highly automated laboratory system that allows a simple blood test to be used to test for nearly 400 allergenic substances. This has enabled it to become the market leader in the field of allergy diagnostics, with annual sales of SKr 1.46 billion representing 70 per cent of the global market.

"There are many similar examples which demonstrate the importance of the close links which we have with the university community," says Ulf Lundkvist, director of research administration in the new company's corporate research and development department.

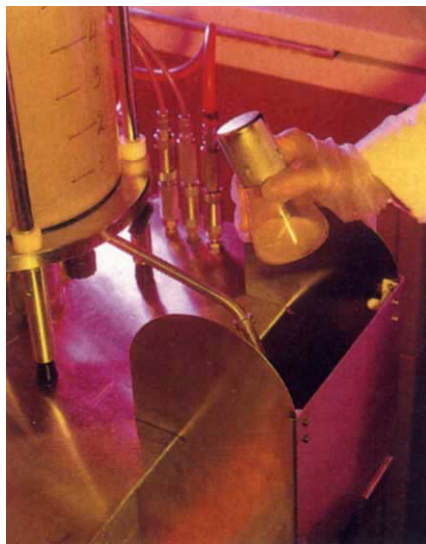
Indeed, the experience of the pharmaceutical industry in Sweden is often

quoted as a paradigm of how an industry can benefit from links to basic research in the country's universities.

What shape these links will take in the future, however, remains to be seen. On the one hand, the new company has decided to align most of its research activities, both in-house and supported in universities, with the development of product lines with which it is already familiar. On the other, it is looking increasingly to research groups outside Sweden to provide the scientific inputs it requires to meet these strategic goals.

Both result from the pressures of international competition, which were themselves one of the main reasons for the merger between the two companies. "We intend to become a research-based international pharmaceutical company, and to have the muscle for long-term research you have to be a certain size; we were too small before," says Ola Rönn, corporate medical director.

The merger has given the joint company a substantial research base; its research and development budget last year was SKr1.4 billion (£140 million), repre-



In-house research is now largely business-oriented.

senting 15 per cent of the company's sales. At the same time, the subsequent rationalization has resulted in the phasing out of what the company describes as "a number of miscellaneous projects and activities in the research area" in favour of a concentration of research on a limited number of strategic areas.

On the basis of successful products already on the market, the company has now decided to concentrate on six such areas: urology and gynaecology; oncology; growth and growth disorders (new insulin-like growth factor, which can be used in situations where there is an

observed resistance to growth hormone); the treatment of blood clots and bleeding; cataract surgery and ophthalmic diseases; and anaesthesia.

Long-term research, especially in molecular biology, is being concentrated in a new Kabi Science Centre, a group-wide research organization based in Stockholm with a special focus on recombinant-DNA technology. This combines the operation of a newly formed Structural Biology Centre and molecular biology research in Kabigen.

The centre provides a central biotechnology resource for the whole of Kabi Pharmacia, specializing in genetic engineering, protein purification, protein engineering and cell culture. Like the rest of the company, it is now organizing much of its research around the six strategic objectives. "About 80 per cent of our research will be directed to the needs of our business units," says the centre's manager, Staffan Josephson.

According to Josephson, the company expects to remain heavily dependent on new ideas and techniques being developed in universities. "But in future, the decision to focus our research strategy means that in some areas we will probably be doing more [with universities], and others we may be doing less."

International competition in the research field will also lead to what Lundkvist describes as a "fundamental change" in the company's relationship with universities in Sweden, namely that they will have to be prepared to compete more than in the past for research contracts with research groups in foreign universities. "Now that we have this international presence focused on a relatively few research areas, we are looking for the best research groups worldwide," he says. Lundkvist points out that the company already has several international collaborations taking place outside Sweden, and expects this trend to continue.

Like many industrial research managers, Lundkvist says he is keen that the government should maintain an active role in helping to bring universities and industry closer together. In particular, the company supports moves by universities to develop internationally recognized centres of excellence. But he emphasizes the need for two-way communication to ensure that these centres respond to genuine industrial need.

"Industry is close to the market and has a lot of R&D experience," says Lundkvist. There is therefore a need, for example, for universities to talk to industry to find out in which fields there will be most the greatest demand for PhDs in the future. "If the universities do not do this, with the free market that we will have in the new Europe, they will just end up educating good scientists who will then go and work in other countries." □

Denmark: Influence beyond its size

DENMARK, perhaps the most optimistic of all the Nordic countries, and certainly the most cheerful, has no inhibitions about its ambitions for the future. A year ago, the Ministry of Education and Research gave it as the government's intention to create a number of research institutions "among the foremost of their kind" and to have "only research institutions of international quality".

There is nothing startling in a small country seeking to have a few outstanding laboratories, of course. The remarkable feature of the Danish claim is that they wish to have none others. Is that possible? People such as Peder Olsen Larsen, chairman of the Danish National Research Foundation, and Knud Larsen, who looks after research for the Ministry of Education and Research, insist that it is.

How can a small country be so ambitious? Danes are quick with easy answers, one of which is historical. Did not Tycho Brahe work in Copenhagen? Then there were a string of more recent innovators beginning with Berzelius and including Niels Bohr. Denmark reckons to have done well in attracting people of daring.

Denmark also has going for it, the research managers say, its habit of collaboration and the much extended scope for collaboration provided by the framework of the European Communities. Danes individually may be apologetic that Denmark fired the first substantial shot across the bows of the Maastricht Treaty last June, and remarkably shy of saying how they voted, but their enthusiasm for the European venture seems

entirely undimmed.

Several at an informal gathering of senior scientists argued that the opportunities for collaboration are now much more interesting than in the years immediately after the Second World War. Then, strictly Nordic collaboration was pushed hard, but the most ambitious and promising graduate students and postdoctoral fellows went off to the United States. Now, people say, European laboratories are often as powerful a magnet.

Denmark also offers the benefits of a conservative and largely rural society, relatively low costs and a polyglot population drilled in speaking English.

Denmark's science is thus still on a rising curve of support from public funds. (Total spending last year was Dkr5.91 billion, including ministerial support for applied research.) Although the total is expected to decline in the three following years, the proportion of total funds directed to basic research of various kinds is likely to become greater.

But a small country cannot do everything. That is why Denmark has set out deliberately to choose priority areas that will concentrate both teaching and research in fields reckoned to be productive. Perhaps inevitably, one of these is biotechnology. Since the beginning of 1991, four of the six research councils (those for natural science, health science, agricultural and veterinary research and technical research) have jointly spent Dkr456 million on research projects at university laboratories and elsewhere.

The expectation is that universities and other organizations participating will

eventually find the costs of the infrastructure they have created from other sources.

The biotechnology programme is claimed to be a model of how to kick-start an area of research. First, there has to be some kind of tradition (food processing and breweries), interested people (the universities were full of them in the mid-1980s), the promise that a local industry would be interested (almost self-evident in Denmark) and the prospect of collaboration with other organizations elsewhere. In the Danish fashion, the question "should there be a biotechnology programme?" was endlessly discussed in advance, notably by the research councils, the universities and the Danish Science Policy Council. From the outset, it was recognized that external evaluation half-way through the programme was an essential. Now, Denmark is slowly silting up with mostly favourable but often bulky evaluation reports from groups of outside experts from Japan, the United States and elsewhere in Europe.

Denmark is by no means a pioneer in the use of foreigners for research evaluation but it is uncommonly comfortable with the procedure. Academics say that they find these reports genuinely helpful, both in drawing attention to the scientific importance of what their projects are about but also politically, by reminding political paymasters that others find their work worth studying. But the practical value of a supportive report is that it can draw attention to deficiencies of funding or other means of support that Danish authorities have heard before but are

Universities look forward to more autonomy

ACADEMICS like to think that even onerous laws do not cramp their style, which is the spirit in which people from the universities of Copenhagen or Aarhus say the 1972 university reforms made little difference to their ways of working. Students may have been empowered to take a part in university government, as were members of the nonteaching staffs and their trades unions. "We know that these things happen", seems to be the message, "but we are reasonable people here and have never had any trouble." That may be the truth, but it can be only be a part of it. Before the early 1970s, Scandinavian countries did not have university legislation, but only universities. The arrangements made uniform in Denmark in 1972 did not merely codify the management of universities, establishing the

rights of departments and faculties against their university at large, but also gave the government powers of direction in matters such as the filling of vacant positions and the continued existence of particular research institutes and departments.

As in much of the rest of Europe, the interval since the early 1970s has been an often unhappy one. At some institutions, the Copenhagen Polytechnic University for example, the decentralization of authority seems to have been a continuing source of friction. Elsewhere, over two decades or so, the universities and the government have found themselves unpalatably at the opposite poles of arguments about vacancies left unfilled. Simple reorganizations within universities have often required too much negotiation without the outside. And the

universities became resentful of the government's pursuit of free access by students without consulting academics.

The result is the new law now widely discussed. Universities will acquire again some of the autonomy that they complain they have lost. As part of the reorganization, groupings together of previously separate departments have been agreed, as have understandings that it will be for the universities themselves to decide which vacant posts to fill. Every institution appears to believe that the future is likely to be better than the past, and even the Copenhagen Institute of Technology — which is the Danish university recruiting uniformly across the country — is uneasy that some of its members may oppose the new law. □

likely to listen to only when they hear them from outside.

But why should there be six research councils to administer such a relatively small budget? Apart from the four listed above, there are the newly created Danish National Research Foundation and a research council for the social sciences. The simple answer is that Denmark likes it that way. The system is designed to allow for endless consultation, which is what the Danish parliament does anyway. Moreover, with six councils in such a small country, the chances are very high that the members of particular research councils will know personally those who are applying for research support, with the advantage that referees' reports can be supplemented by personal knowledge. And of course it is always helpful to have more than just one chief source of funds.

Even so, Denmark is lucky to have avoided the difficulties that arise when, in a small community, grant-making consists of spreading what is available uniformly. The trick seems to have been not to be afraid of speaking out. That is helped in two ways — by the fact that every Danish academic can look to his university for basic research support and because most Danes seem uncommonly good-humoured. □

New name for an old friend

One of Copenhagen's monuments that will survive is the facade of the Niels Bohr Institute, now almost hidden and certainly dwarfed by newer buildings of the University of Copenhagen as a whole. But the institute will, from next year, change its name and will become the Niels Bohr Institute of Physics, Astrophysics and Geophysics. Reverence for Bohr apart, the institute has been forced to recognize, during several lean years when academic vacancies could not be filled, that safety for the future rests on the construction of a larger academic unit, able to make a more obvious claim that it is active in both teaching and research. As a part of the deal, the government has apparently promised that it will never again be asked to leave academic posts vacant.

In reality, the Niels Bohr Institute remains as much a mecca for graduate students and postdoctoral fellows from abroad as it has ever

WHAT does a government do with the proceeds of selling (or privatizing, as the British say) nationally owned assets? Sometimes the money simply disappears into the national treasury. In Denmark, the sale last year of a group of publicly owned insurance companies has led to the formation of an entirely new research council, the Danish National Research Foundation, with a capital of DKr2 billion. It will be Denmark's sixth and perhaps most controversial.

But why should anybody object to another pot of money for supporting research? There are several answers. One is that the existing research councils are not that well supplied with funds, even as things are. Another is that the terms of reference of the new council require it to make grants not to individual researchers but to groups of them or even, sometimes, to institutions: it is a fair question whether research projects of that kind deserve an estimated DKr150 million a year. And some complain that the new council will support research across the whole spectrum of scholarship, humanities as well as science and technology.

Professor Peder Olsen Larsen, the Aarhus physicist who is the first director of the new foundation, believes that the academic community will be persuaded when the first year's allocations of grants are made. So far, 350 research proposals have been screened by expert committees and individual referees. Applications on a short-list of 50 are now being discussed in detail with the applicants, who may have to make precise

been. Part of the explanation for that is that it lives cheek by jowl with the Nordic research enterprise NORDITA, which accommodates a necessarily more transient group of scientists with interests in

previously vague budget estimates and will eventually be sent more formally for review (as often outside Denmark as within it). The first awards may be announced in the spring.

Some lucky survivors to the short-list say that the new council may be backing away from its first suggestion that grant-holders should not simultaneously be members of a university or some other institution. The potential benefit of the new council is the flexibility it will have in apportioning funds between salaries, equipment and running costs and the chance that some of the grants it makes may be several millions of kroner.

Academics say that the value of these schemes is that they may allow people quickly to establish substantial groups of researchers in important fields. In recent years, the targeted biotechnology programme, which enabled a number of universities (Aarhus is one) to establish important research centres, has shown the potential value of large grants spread over several years.

But the new research council is meant to cover the whole of scholarship and is required by its enabling legislation to regard the erosion of its DKr2 billion endowment by inflation as a first charge on its annual income. The result of that may be that, with the continuing uncertainty about the interrelationship of European currencies, the research council may have little to spend in some years. It may be just as well that the council has let its first year go by without spending anything substantial on research. □

ships but who are otherwise indistinguishable. NORDITA has itself been looked at carefully in the past few years, as part of the Nordic countries' repeated external evaluation of their institutions.

On the basis of what the external critics say, it seems likely that NORDITA will itself shift to a more distinctive pattern of research with particular topics being chosen in advance to be the subject of the institute's research for periods of three or six months at a time. But the same international evaluation of NORDITA took the view that it would not be proper to go so far as the Institute of Theoretical Physics at Santa Barbara, California, where there are funds sufficient to ensure that the institute's management can concentrate resources on the fields that it has chosen. □



The Niels Bohr Institute — still as much a mecca as it always has been.

particle physics, nuclear physics and astrophysics, who are on short-term fellow-

Aarhus on the accelerator

AARHUS (correctly spelled Århus) is an agricultural market town in the western province of Denmark which is dominated by a new, quickly growing and excellent university. If you arrange to get there by air and early in the morning, at this time of the year the place will look like a Japanese print, with the tops of trees and occasional red brick and tile houses protruding from the patches of uncleared mist.

The presence of the university in this far-off place is yet another proof of the eagerness of Scandinavians to be regionally self-contained. Aarhus began in the 1920s as an agricultural school to which a medical school was added in the mid-1930s. Only then did the founders appear fully to appreciate medicine's need of service teaching; soon after the end of the German Occupation, a group of physicists moved out from the Niels Bohr Institute in Copenhagen, for example, to found what has now become the outstanding school of atomic physics.

The regional character of the university is, of course, embodied in the student body, not the faculty (which is as mobile, within Denmark at least, as anybody could ask). As things are, the total student population is pushing 28,000 and, as elsewhere in Denmark, all qualified school-leavers have a right to be taught, usually in the faculty of their choice. Most of them, at Aarhus, are from the western region.

The vice-chancellor of the university, diverted the other day from his entertainment of a group of US academics on a grand tour of European universities, had

no doubt of the reasons why it has been possible for a regional university to win an international reputation, at least in some fields. He explains that the academics who teach at Aarhus know that they must choose between rustication and the maintenance of an international connection — and that he does everything he can to ensure the appointment of people in the second category.

That seems to explain the rapid growth of the physics department's reputation since the early 1950s. Its founders did not sever their connections with the Niels Bohr Institute when they moved from Copenhagen. On the contrary, they remained the driving force for a Danish scheme to build a kind of universal heavy ion accelerator (a cyclotron for accelerating heavy ions in which the trick was to develop a sufficiently broad range of ion sources to cover all eventualities).

In the event, the machine itself was never built, but parts of the technology developed for it are still used in the programme of which Aarhus has made a specialty of its own — the study of collisions of fast negative heavy ions. (Negative ions, with extra electrons, have large cross-sections with the components of the residual vacuum in any machine, and are in any case intrinsically less stable to electrical perturbations.)

The tangible embodiment of this ambition is a unique storage ring for heavy ions (called ASTRID, for "Aarhus Storage Ring in Denmark"). The storage ring, reminiscent of GeV accelerator equipment, but on a human scale, is tucked neatly into a single ground-floor laboratory. And the ring is not circular

but square — four 10-metre sections joined at the corners by bending magnets.

The trick with such an apparatus is to operate at a vacuum good enough to give stored ions a reasonable lifetime against collision. The Aarhus group reckons to be able to keep negative ions in circulating form for periods of the order of several seconds. Only thus is it possible to make meaningful measurements of the intrinsically unstable products of, say, the excitation of a stored ion.

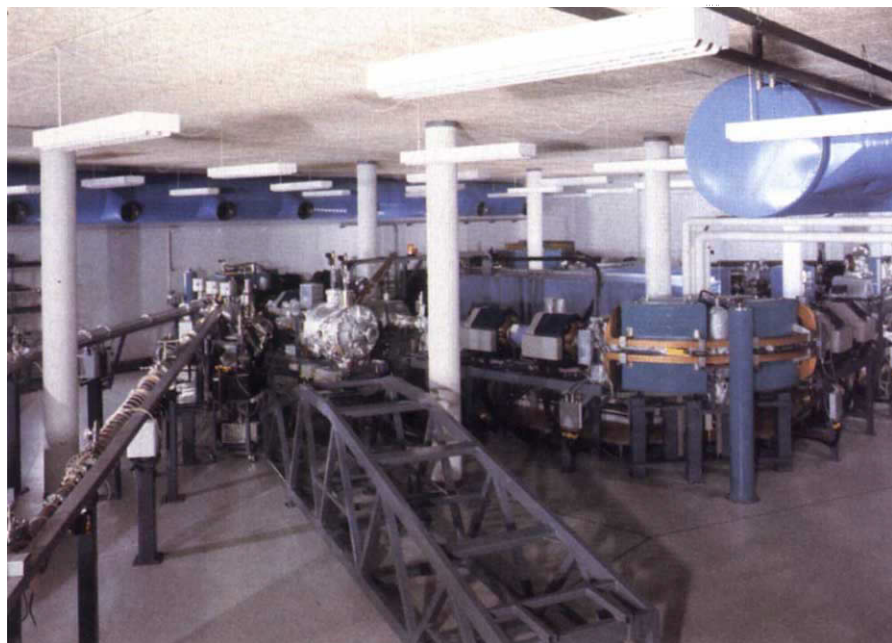
That same development is part of the reason why the physics laboratory now boasts that more than a dozen cyclotrons of various kinds are used as parts of specialized experiments.

This is not, of course, high-energy physics. (The department cherishes its connections with CERN, the European Laboratory for Particle Physics at Geneva, especially on the theoretical side, instead.) But it is a way of gathering data about previously unstudied atomic constructs, the spectroscopy of the negative helium ion for example (the atom with three electrons and a nuclear charge of two whose spectroscopy should resemble that of neutral sodium).

Indeed, the techniques at which the Aarhus physics department now reckons to excel — not merely the acceleration of negative ions but the trapping of charged particles in electromagnetic traps of various kinds and the use of high-precision laser spectroscopy — were those that made possible four years ago a unique measurement of the two-way velocity of light which remains the most precise estimate yet of this quantity.

The physics department would like to be known for its high output of articles in distinguished journals, but it knows that the bread and butter of its success is the degree to which it is generally accepted as a partner whenever like-minded people elsewhere are engaged on projects of common interest. If you can always place a newly graduated PhD student at a good laboratory overseas, or find a ready welcome for a graduate student looking for a few months experience with somebody else's technique, you have to have a reputation and you have to be able to give as much as you get. But if those conditions are satisfied, it does not matter that you are placed in one out-of-the-way region of one of Europe's smallest peripheral countries.

The cell biology department in the faculty of medicine at Aarhus, being newer, works the international connections differently. Evidently the European Communities' contribution to the Human Genome Project has been a godsend, providing the department with a contract to investigate the tissue-specific distribution of human proteins by two-dimensional electrophoresis. The techni-



The 'square-ringed' ASTRID device.

Fertile grounds for the growth of biotechnology

THE Carlsberg Brewery in Copenhagen is a sprawling industrial estate at the centre of which is a heavy redstone entrance to what had been the original plant, now cocooned as a kind of museum. Modern Carlsberg is spread across the world in more than 40 countries, the more so since Carlsberg swallowed, a decade ago, Tuborg, Denmark's other big beer. But the brewery, an early manifestation of biotechnology, is now doing a great deal to return the benefits it has derived over centuries from science.

The Carlsberg Foundation is not simply a shareholder in the brewery but is also a research organization in its own right. Like other research foundations with roots in a science-based company, the foundation makes grants for good works — scientific meetings, Danish researchers' travels abroad and so on. But now it is also directly engaged in basic research on the brewery site. And because it seems to the pragmatic men in charge to make no sense to have two scientific organizations with similar ownership and objectives working separately in the same

place, relationships with the works laboratory are strong and getting stronger.

Klaus Bock, head of the chemistry division at the Carlsberg Foundation laboratory, explains that the requirement that his research programme should have something to do with the basic science of brewing is, these days, not an irksome restriction. For one thing, most organic



Definitely Carlsberg's oldest brewery ...

processes resemble the transformations of the brewing process in one way or another. But yeasts, in any case, are useful model organisms for all kinds of studies.

For the past ten years, the bulk of the foundation's research has indeed been biotechnology. Bock reckons that the Carlsberg Foundation has done as much as most other organizations to introduce these novel techniques to Denmark as well as to stimulate the government's own interest in an earmarked biotechnology programme.

But that, Bock says, does not mean that the foundation's interests will be swamped by work stimulated elsewhere. The foundation's research park continues to take in students (from universities all over Denmark). It is also a productive centre of post-doctoral fellowships and has sufficiently deep pockets to attract people away from other posts when it feels it needs them.

But why does Denmark so emphasize biotechnology? Bock has a simple answer. Apart from brewing, the traditional industries,

mostly turning milk into cheese, made people worry about germs even before Pasteur introduced the term. Now, the techniques of fermenting and maturing hygienically are in the Danish blood. □

ques (and even the software) used for the analysis of data are familiar in a dozen laboratories worldwide. The calculation is that successful management of the European Communities contract will enable the laboratory at Aarhus to rub shoulders with the others in the field. This is how to bootstrap your way to international status — but with the proviso that you must succeed with the contract that may give you that right.

Both the physicists and the biologists of the medical faculty agree about the general problems of doing science competitively in small countries. Surprisingly (for this is almost unprecedented talk), everybody agrees that there is no real shortage of funds for travel. Faculty members can attend meetings they need to without much difficulty and the small amount of bureaucracy with which they must contend is probably just that needed to ensure that they do not attend peripheral meetings as well. But even grants for postdoctoral fellowships and the departmental graduate education ensure that unestablished people are also able to travel. One need of small-country science is evidently that its public sponsors should recognize the need for physical

movement.

The people at Aarhus also congratulate themselves that, unlike many in the United States, for example, they are not required to compete for funds for technical assistance. The university budget allows for a hard core of technical staff to be kept on the university books and to become the kind of technicians who are able to contribute as much to the development of technique as do researchers themselves.

But not every department in a university such as Aarhus can realistically hope to be up with the best. Indeed, one of the crucial problems of countries such as Denmark (mirrored it appears in the other Nordic countries) is the way in which essentially open access by students forces the need for a faculty which is itself comprehensive — one in which many teaching loads will be less equal than others. It is agreed that only discrimination in the allotment of resources between one department and another will allow this task to be done well.

The dean of science at Aarhus, Professor Carl Petersen, says openly (even in the presence of his dependent heads of department) that even a large university

in a small country cannot succeed unless somebody has the authority to make clear decisions.

So much will be assured by the new University Law, now making its way through the complicated consultative procedures of the Danish parliament. In retrospect, it seems agreed that the legislation of the early 1970s formally went too far to invest power and authority in students and nonteaching staff. The new arrangements will take back some of those concessions but also require the universities to reorganize their teaching the better to suit the needs of their students. Under the same legislation, the heads of research institutes, deans and rectors will have more power than before.

Formally, that means that the capacity to make clear decisions should be confirmed. It also, at Aarhus, seems to mean that most departments will take the chance to put graduate courses on a more formal basis than at present, partly replacing the standard research project with more formal teaching. Some regret that prospect. Others, like Carl Petersen, say "there's no other way of getting the throughput we need". □

Norway: It's time to reorganize

SOME call it a brave experiment in science policy; others a foolhardy move doomed to failure. Either way, Norway will next month take a step that several countries — including Britain — have contemplated but so far backed away from: the creation of a single funding council covering all fields of both basic and mission-oriented research, from the study of Nordic myths to the applications of semiconductors.

The experiment is designed to bring greater coordination to the country's research efforts, in response to long-standing criticism that these are too fragmented. The government argues that a single research council will make both policy-making and research funding more effective. Its critics reply that any attempt to centralize decision-making risks ignoring critical distinctions — such as the different criteria needed to evaluate basic and applied research grant applications — and may eventually prove unworkable.

At present, 25 per cent of all public funding of research and development in Norway, an estimated total of Nkr 2.22 billion (£225 million) this year, is distributed by five different research councils. Two are funded through the Ministry of Education, Research and the Church. The largest, with a current budget of Nkr613 million, is the Norwegian Research Council for Science and the Humanities (NAVF), established soon after the Second World War as the main channel for the support of basic research in universities and research institutes. The second is the Norwegian Research Council for Applied Social Science (NORAS), with a budget of Nkr174 million.

The three other research councils are each funded through the government ministries responsible for their field of operation. The most important of these is the powerful Royal Norwegian Council for Scientific and Industrial Research (NTNF), also set up in the 1940s. Its budget of just over Nkr1 billion in 1990, until now paid by the Ministry of Industry, represents 47 per cent of Norway's total research council expenditure.

In addition, the Norwegian Fisheries Research Council (NFFR) and the Agricultural Research Council of Norway (NLVF) account for 7 and 8 per cent of the research council expenditure, receiving their funds through the ministries of fisheries and agriculture respectively.

Demands for greater coordination of Norway's research activities — and in particular closer links between the research council — have been voiced in a series of national and international assessments of Norwegian science policy

over the past 20 years. Up to now, however, such proposals have met strong resistance from the individual ministries, keen to maintain direct control over research programmes intended to meet their sectoral needs.

The deadlock, however, has been broken. Centralized funding arrangements were proposed in a report prepared for the government last year by a committee headed by Knut Grøholt, a senior civil servant. The committee's main recommendations have been accepted by the government with the personal backing of the Prime Minister, Mrs Gro Harlem Brundtland, who herself carried out research before embarking on a political career. They are now being actively promoted by the Minister for Science, Education and the Church, Gudmund Hernes, who is also a professor of sociology at the University of Bergen.

Under the new arrangements, all research council funding will be channelled through a single council, the Norwegian Research Council (NRF). This will be directly responsible to Hernes, and run by a single administration within his ministry. Research funds from other government departments, which would previously have gone directly to the individual research councils, will now be channelled through the NRF to one of six sub-councils covering the following fields: industry and energy; medicine and health; environment and development;

In line with the conclusion of the Grøholt committee, the government hopes the new arrangements will encourage more cross-fertilization between the academic community and Norwegian industry. "At present, there is little cooperation between the councils, meaning that it is very difficult to have an overview of what is going in and what is coming out," says Ole Didrik Lørum, rector of the University of Bergen, and chairman of the interim board set up pending the board's official creation next month.

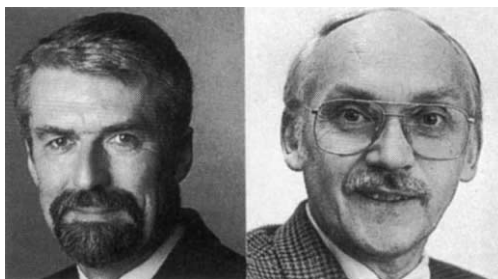
Rolf Skår, director general of the NTNF, is enthusiastic about an integrated approach. "We have been very much in favour of this change since it was first announced, for the simple reason that we feel it is very bad science policy to have a separation between fundamental and applied research" says Skår. He adds that he is particularly keen to see the work of universities and of the research institutes linked in more directly to the needs expressed by Norwegian industry.

How the arrangements will work in practice remains to be seen. The 10 members of the new research council are taken not only from the research community, but also from government agencies, private industry and the trade unions. This, some fear, could create problems when it comes to the politically sensitive issue of how funds should be divided up between different fields and different types of research.

Others have pointed out the potential difficulties of bridging two very different missions and funding cultures. "I am fairly sceptical about the change," says Hans Skoie, professor of research policy at the Institute for Studies in Research and Higher Education in Oslo. "The NTNF is really more like an industrial agency than a research council. In all other Nordic countries, such agencies are under political control; in Norway, this type of applied and mission-oriented agency is now being merged with a classical university-oriented research council which is part of a dual funding system operating on very different values."

Bjarne Waaler, professor of physiology at the University of Oslo and president of the Norwegian Academy of Science and Letters, is concerned that basic science may suffer in a drive to emphasize its industrial applications. "You cannot coordinate every type of research activity in a well-planned way, both vertically and horizontally; that is overambitious."

Such fears are dismissed by Lørum. "The best way to have good applied



Two supporters of the plan to integrate the research council's — NTNF's Rolf Skår (left) and Didrik Lørum of the University of Bergen.

bioproduction and processing; natural science and engineering; and culture and society.

In addition to providing a channel for public funding of research, the new board will have responsibility for advising the government on science policy issues, filling a vacuum that has existed ever since the abolition of the country's Research Policy Council in 1988. It will also serve as an umbrella organization for coordinating the research carried out in the country's many publicly funded research institutes.

research is to have strong basic research, so I strongly support the idea that the two should be integrated," he says. Lærum defends the government's decision to appoint individuals representing a wide range of social interests to the board of the new council. "The thinking behind this is that the research councils should serve society, and that the make-up of the members of the board should therefore reflect society."

But even Lærum admits that there are still many details to be resolved before the new arrangements are put into place — for example, how much control other ministries are likely to be able to exert over the individual sub-councils. Meanwhile, the academic community is reserving its judgement: while the idea of a more streamlined system has been generally welcomed, approval is mixed with caution that only time will dissolve. □

Norway's universities challenged

NORWAY'S universities are bursting at the seams. The government's simultaneous desire to increase the knowledge and skills of its population, and at the same time keep unemployment levels as low as possible, has encouraged it to support a massive expansion of undergraduate places, which rose from 102,000 in 1988 to 140,000 in 1991. Each of Norway's four universities — in Oslo, Bergen, Trondheim and Tromsø — has grown rapidly over the past decade, as have its six university-level colleges.

The explosion in demand has been a mixed blessing for the universities themselves — and in particular for university scientists. Many of the new jobs have gone to create new positions in science and engineering faculties. According to official statistics, the research budgets of the universities have increased substantially as a result, as half of the salaries of university scientists are supposed to cover the time they spend on research. But often the additional funds needed in practice to carry out such research have not materialized.

Pressure on resources resulting from the rapid growth has coincided with an increased concern over ways of raising the quality of research output. These two trends have encouraged the government, despite being dominated by the Labour Party, to modify some of the more egalitarian principles that have dominated the distribution of research funds in the past.

In particular, the general principle that all university scientists receive an equitable share of a university's research money is slowly giving way to increased selectivity in the allocation of resources. "We have now decided that in universities, there should be more money for the more productive groups," says Tore Olsen of the Ministry of Education, Science and the Church. "For example, if a particular university research group produces more PhD students, we have decided that the group should get a larger share of the money going to that university."

In a similar vein, the Norwegian Research Council for Science and the Humanities (NAVF) has, for the past few

years, been allocating block grants to research. Both developments represent a significant shift from egalitarian to elitist principles. "That movement has much support from the government, and I think that the support is widespread in the community at large," says Olsen.

The universities themselves have agreed to explore ways of creating "centres of competence", accepting a mutual division of labour between institutions. These centres will be linked together as the nodes of an integrated higher education system, dubbed 'Network Norway'.

Closer administrative ties between the universities will be backed up by a complex system of electronic linkages and modern data communications technology.

On the political front, the universities, which have traditionally been fiercely independent of each other, have now joined forces to create a new body, the Norwegian Council of Universities. "We have achieved far better coordination than existed four years ago, and having one university council will make it easier to deal with one research council", says Ole Didrik Lærum, professor of medicine and rector of the University of Trondheim, who sits on both bodies and provides a link between them.

Many university scientists have welcomed the concentration of funding. "We are not rolling in money, but we are certainly better off than we were before", says Pers Anderson, professor of brain research at the University of Oslo, and one of the recipients of extra funding under the new block grants scheme. "Already we see a flow of activities, such as being able to send young researchers to international meetings, that we would not have been able to achieve with the new money." □

Resisting the northern drift

To move or not to move? That is the dilemma facing the Norwegian government over the fate of the Polar Institute in Oslo, one of the country's leading research institutions for Arctic research and a key participant in current debates over the fate of the world's polar regions.

The institute has been located in the capital since it was founded near the beginning of the century. As on numerous occasions since then, the government is under strong pressure to move the institute up north, possibly to the city of Tromsø.

Not surprisingly, the University of Tromsø (see next page) has been an enthusiastic advocate of the institute's suggested move, arguing that the Polar Institute would form a natural part of its efforts to become a national and international centre for Arctic research.

Resistance to any move is, however, strong among the institute's scientists. According to a recent survey, fewer than five out of a total staff of one hundred said they would be prepared to move up north. Job and housing shortages in the capital that had previously justified the government's efforts to decentralize its activities no longer exist in a period of rising unemployment and falling house values. "We disagree strongly with the proposal that we should move up to

Tromsø, and are not prepared to go along with it", says the institute's director, Orits Land.

Tore Vorren, chairman of the Roald Amundsen Centre for Arctic Research at the University of Tromsø, disputes the argument that Tromsø is too far from Oslo for effective interaction with government officials. "We have come a long way with telecommunications, and do not think that there would be a problem with that."

According to Land, however, the key role played by the institute in advising government departments on Arctic policy makes it essential that it remains physically close to the centre of power. It is also important, he says, that the research activities of the institute should not — as some have proposed — be split from the advisory role that it provides to government departments. "We feel that our advice to government should be directly based on our practical experience," he says.

The government itself is divided on the issue. Departments that rely heavily on advice from the institute want it to stay in Oslo. Others, more sensitive to the pressure of regional demands, favour devolution. A final decision, which many now believe will indeed involve splitting the institute into two, is expected in the next few weeks. □

The view from the North

"SOME people are astonished to find that there is intellectual life north of the Arctic circle", says Are Johnsen, director of information at the University of Tromsø on Norway's northwest coast. The university was established in 1968 in a deliberate effort to develop the economic and social infrastructure of northern Norway. Situated at a latitude of 70°, Tromsø prides itself on being the most northerly university in the world.

Tromsø's distance of more than 1,000 km from Oslo, together with its relatively inhospitable climate — the snow that fell in October this year is likely to remain until next May, and the university campus does not see the Sun for more than two months in the winter — has not always made it easy to recruit either students or staff since the university was opened by King Olav V in 1972.

But with its academic record growing, and national unemployment now at record levels, students are queuing at the door. The university has increased rapidly in size over the past four years, and its current student body of more than 5,000 is almost twice the size originally predicted.

Recruiting research staff has been made easier by some of the special conditions that the university has been able to offer. For example, all faculty members are given 12 months sabbatical leave every four years, and travel grants are more generous than at other Norwegian universities. Both have helped the university to build to up its current strength of 400 research workers, working on 700 projects.

Many of these relate to the specific conditions of the Arctic region. Tromsø has long been both a trading post and last port of call for fishing and exploration expeditions alike, and is proud of its reputation as the gateway to the Arctic. Its scientific importance has grown steadily as interest in the region's climate and environment has increased over the past decade, and the university has been keen to develop a role as "the university of the Arctic".

Much of this research is carried out under the aegis of the Roald Amundsen Centre for Arctic Research, named after the explorer who had close connections to the town. To celebrate the centenary of a famous exploit by the Arctic explorer Fridtjof Nansen, the centre will play a key role in plans drawn up by the Research Council for Science and the Humanities (NAVF) to mount a scientific expedition in 1994 that will, like Nansen's three-year voyage in 1893–96, involve freezing a research vessel into the ice and letting it drift across the Arctic Ocean.

The university has long been a centre for research into the biology of animals living in the Arctic region, and is in the process of

building a new Nkr 40 million (£4 million) research centre for its department of arctic biology, with funds from four separate government departments.

Physics is another field in which the university has been able to exploit its unique geographical situation. Tromsø's proximity to the Gulf Stream has made it a relatively comfortable location from which to observe the aurora borealis. The university's physics department has been built around an observatory set up by the meteorologist Kristian Birkeland in 1928.

The department's interest has grown to include fields such as optical astrophysics and molecular physics. But with pressure on all universities to focus on "centres of competence", it is now concentrating on building up its expertise in plasma physics.

"We would like a centre that has the capability to study the physics of the atmosphere from the ground upwards", says Asgeir Brekke, professor of physics, who says that he hopes the department will be selected as one of the nodes in the 'Network Norway' currently being established by the country's universities (see page 521).

The university is keen to use its research capacity to expand its role in regional affairs. For example, it is already preparing a report on the way in which joint research programmes carried out with scientific teams from the region of Russia which joins on to Norway's north-east border could help to cement political stability in the region. Potential topics for joint study include regional employment and the potential environmental effects of reopening the North-East Passage to Japan.

Looking in the opposite direction, the university intends to make the most of the

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Meteorologist Kristian Birkeland established an aurora laboratory in 1928, the cornerstone of Tromsø's physics department.

possibilities being opened up by Norway's likely membership of the European Commission's Framework Programme. Tore Vorren, director of the Amundsen centre, is quick to acknowledge the irony of the situation, as the local region is strongly opposed to Norway's application to join the European Communities.

"Two thirds of the local population are against such a move," he says. "But as scientists, we look upon this as an opportunity, and a chance to get in better contact with the European scientific community. This would certainly be a good thing, since at present many of us have better contact with the United States than with scientists elsewhere in Europe." □

EISCAT's place out of the Sun

NORWAY may have relatively few major scientific achievements to its name. But it does offer a number of unique locations for scientific experiments and observation. One is the Svalbard archipelago, half-way between Norway's northern coast and the North Pole, which has just been adopted as the site of a new radar to be added to the European Incoherent Scatter (EISCAT) system.

The new radar will initially include a 500-kW transmitter and a 32-metre diameter receiver. Provisional approval was given by the EISCAT council last month. The final outcome now awaits a decision from Britain's Science and

Engineering Research Council, which is being asked to provide 18 per cent of the construction costs of SKr106 million (£10.6 million).

EISCAT was established in 1977 as a collaborative effort by European scientists to investigate the interaction between the particles in the solar wind and the Earth's magnetic field near to the pole (the origin, for example, of the aurora borealis). With its present facilities, radio waves are transmitted into the ionosphere from Tromsø, and the scattered signals are subsequently picked up at receiving sites at Sodankyla in Finland, at Tromsø itself, and Kiruna in Sweden,

which is also the location of EISCAT's administrative headquarters.

Under the terms of the original agreement signed by six member governments, 25 per cent of the operating costs are currently carried by its three largest member states — Britain, France and Germany — and the remaining costs are split between Norway, Sweden and Finland in a ratio 10:10:5.

The presence of the EISCAT transmitter has certainly helped to put Norway on the scientific map at a relatively low cost. "If you look at the economic side alone, it is a fantastic scoop for Norwegian science," says Asgeir Brekke, professor of physics at the University of Tromsø, who was last month elected chairman of the EISCAT council for the next two years.

"We pay only 10 per cent of the running costs, and the rest is paid for by the other participating countries. The result is that we are bringing in more money to the Norwegian community than we are spending. That is very unusual — to say that you are doing basic research without being a load on the taxpayer."

EISCAT itself has already proved to be a highly successful facility, producing a string of useful scientific results on the physics of the upper atmosphere. Svalbard's position means that the new radar, which will be operated in co-ordination with the existing facilities, will offer space physicists a unique opportunity to observe those regions of the ionosphere and magnetosphere that are currently inaccessible to ground observation, including the region immediately above the North Pole.

In particular, the new radar will allow the simultaneous observation of the interaction of solar-wind particles with the ionosphere by optical and radar means, since during the winter months it is still dark during the day when particles from the Sun have relatively direct access to the atmosphere.

Norway's enthusiasm to see the new radar built has led it to agree to contribute SKr37 million towards the construction costs, a considerably higher percentage than its current contribution to EISCAT. This is partly explained by the fact that the Norwegian government's enthusiasm for the new radar is not motivated just by the science.

The construction work on the radar — and subsequently providing the infrastructure and services needed to maintain it — will provide much-needed jobs for a region now experiencing high unemployment as a result of a decline in the local mining industry, previously the principal source of work. And continued political instability in Russia means that Norway is keen to maintain a presence in the region. □

NATURE · VOL 360 · 10 DECEMBER 1992

Alternative view on whaling

WHEN Mrs Gro Harlem Brundtland, the Norwegian Prime Minister, chaired the international commission that endorsed the term "sustainable development" in the mid-1980s, she became a heroine of the world's environmentalists. This summer, however, her halo became severely tarnished when Norway announced that it planned to defy the current international moratorium on commercial whaling.

To many outside Norway, there seems a clear conflict between a desire to respect nature and a desire to exploit it. But inside Norway, the positions are seen as entirely compatible. "Any resource which can be harvested on a sustainable basis — and as a scientist I believe that this applies to whales — should be harvested," says Arnoldus Schytte Blix, director of the department of arctic biology at the University of Tromsø. "Not doing that is a waste of resources."

Inevitably, the country's scientific community has become embroiled in the controversy over Norway's stance on commercial whaling. One reason is the government's desire to demonstrate that its position is supported by scientific evidence that whale populations are not in a particularly precarious state (just as it has used comparable arguments elsewhere to justify restrictions on exploiting various species of fish).

"The most important part of our research on whales has been the nonlethal part, in particular the development of new methods for counting whale populations," says Lars Walløe, professor of physiology at the University of Oslo, chairman of the country's Natural Environment Research Council, and a close adviser to Mrs Brundtland.

In particular, Walløe points out that the scientific committee of the International Whaling Commission has now accepted Norway's assessment that the world's population of minke whales is about 80,000 — not 18,000, as had previously been estimated.

There is still dispute over the algorithm being used to calculate the precise number of minke whales that can be caught without posing a serious threat to the remaining population. More controversial, however, is the fact that Norway has also announced that it intends to use 400 of the whales caught over the next three years to study their feeding habits.

Walløe defends this research as necessary for the future not just of the whaling industry but of Norway's fishing industry more broadly. "We know that minke whales eat a large amount of food; they may eat crustaceans, they may even eat cod," he says. "Finding out the eating

habits of whales is not perhaps a major scientific goal in itself; but from the fishing industry's point of view it is fairly important."

Blix points out that analysis of 50 whales studied over the past five years during which "scientific whaling" has been permitted under the terms of the international moratorium has already produced detailed results on how much energy an individual whale needs at different times of the day and year.

"The big missing link is their diet: what do they choose to eat, and in which part of the Atlantic?" says Blix. "You need to know more than just the number of individuals involved; you need to know where they are and what they prey on at different times of the year."

Blix is quick to respond to charges that he is now seeking an excessively large number of whales for the second part of



Norwegian Prime Minister Brundtland, both heroine and villain in 'green' circles.

the study. "To arrive at data on food consumption, on a species-wide basis, we need to look at animals in five different areas of the North Atlantic and the Barents Sea at three different times of the year; taking, for example, six whales in each sample already gives us a total of about 100 whales a year."

Some scientists feel that the government's decision to allocate the relatively generous sum of Nkr120 million (£20 million) over a five year period to whale research represents an unacceptable political bias. They argue that the government appears more concerned with legitimizing its position in international negotiations and in responding to the demands of Norway's powerful whaling lobby than in meeting genuine scientific curiosity.

"Political motivation is stretching too far into the science," says Dat von Graven, a biologist at the University of Trondheim. Others are beginning to express the views of animals rights groups in other countries that whales should be given special protection from whalers and research scientists alike.

Blix, however, is unfazed by criticism. "Hunting, trapping and fishing have been part of everyday life up here; until recently, they have been activities on which your life depended. In the near future, Norway will be the only Western country in which you can maintain a rational link with nature." □

Finland: Faith in a European future

FINNISH science is emerging from its sauna. The past two decades have seen a faster growth in the country's research efforts, in both public and privately funded research, than in virtually any other member state of the Organisation for Economic Co-operation and Development (OECD). But the economy was overheating. Today, the party has come to an end, and the immediate task — both for the country and for the scientific community — is to cope with the hangover.

The new situation is already threatening the research community. Equipment budgets are being cut and expansion plans postponed. In the public sector, the salaries of scientists — like those of all employees — are being reduced by 4 per cent next year, and job losses are also threatened. In private industry, long-term research is being abandoned in favour of short-term profitability.

It is all very different from the situation only two years ago, when the research budget was still climbing sharply, having virtually doubled in real terms over each of the previous two decades. "It has been a wonderful time for research; everything has been going up", says Elisabeth Helander, research director at the Academy of Finland, the body responsible for distributing funds for basic research to the universities through its seven separate research councils.

Helander points out that the extra funding has produced significant improvement in both the quantity and the quality of research produced by Finnish scientists. For example, there have been increases both in the number of papers published in international scientific journals and in their impact on the scientific community. The universities have also achieved a substantial increase in the production of PhDs.

But basic science in general is now facing an uncertain future. The government announced in September that it wants to cut the academy's budget for next year by 5 per cent from this year's total. Taking into account inflation, the real decrease in funding will be even higher. The government argues that like the cold shower which follows a good sauna, the current crisis will prove to be an invigorating experience from which Finnish science can emerge leaner and fitter.

Those less optimistic, however, warn that inadequate funding for research and an excessive belief in the pull of the

market, rather than in a government-led innovation policy — for example, in fields such as microelectronics and biotechnology — threatens to deprive the country of the fruits of its heavy investment in research in recent years.

Finland's current suffering results from the joint effects of the collapse of its trade with the Soviet Union and the economic recession in Western Europe. The gross national product (GNP) fell by 6 per cent last year, and unemployment is currently at 14 per cent. The centre-right coalition government that came to power last year has responded by taking a knife to the welfare society that has been built up since the Second World War.

Science has been a beneficiary of these past policies. Successive governments have channelled funds into the university sector and a wide range of publicly-funded research institutes. Funding for research and development grew from 1.2 per cent of GNP in 1981 to 2.0 per cent last year — a faster rate of increase than in any other OECD country.

This trend is now going into reverse. "Most of our scientific equipment is imported, and as a result of the devaluations,

up, people's spirits went up too. But scientists are now worried," says Tanskanen. "They think that the money will be taken away from basic science, since politicians feel they should support re-search which promises a quick return. But we must not cut down on our investment in human capital; the people trained in universities will move into applied research in industry, and that is what the country needs."

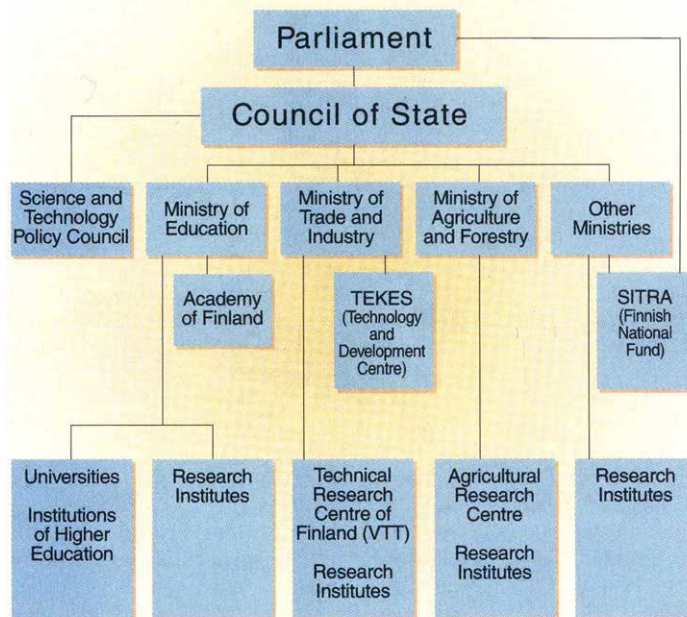
The academy is already responding to the new economic and political climate by looking at a possible restructuring. As part of this process, the Ministry of Education has commissioned a complete reassessment of the academy's activities, with the aid of an international team of evaluators, to see the extent to which its funding strategies should be changed. In the past, for example, the academy has concentrated its research funding on proposals received from individual scientists. Now it is looking at ways in which it can organize researchers into larger groups, increasing multidisciplinary cooperation and focusing its support on areas of particular scientific strength.

The task of re-evaluating past strategies in the light of the new economic climate is also occupying the Council for Science and Technology Policy, the government's principal policy advisory committee on science and technology. The council is chaired by the prime minister, and brings together senior representatives of government departments, research institutes, the academic community, trade unions and employers.

One of its main tasks of the council is to set out broad policy guidelines to provide a framework for how research funds are spent. Two years ago, when the science budget was still increasing, one priority was to establish a "national system of innovation", which would incorporate policies for promoting education, training and research.

The council's main concern now is how to preserve this initiative in the light of the government's attempts to escape its economic difficulties. "The problem that we face is whether we can incorporate longer-term structural aspects into decisions about where cuts [in public spending] are going to be made", says Erkki Ormala, chief planning officer to the council. "The question is, will

How Finland runs its science



the cost of such equipment is going up very rapidly," says the academy's president, economist Antti Tanskanen. "It is too early to say exactly what the size of the cut will be for next year, since the government has not yet agreed on a final figure, but it will certainly have a significant impact."

The mood of the scientific community is already changing. "When the money went

we be able to preserve the objectives of science and technology policy, such as improving the quality and internationalization of research; or are we just looking at conventional questions of regional balance and educational needs?"

Ormalá argues that supporting innovation need not be politically motivated. "What we are trying to create is not *dirigisme*" he says. "It is politically neutral, but it is not indifferent on the question of industrial renewal — which means that we need more R&D, more education and training, and more risk financing."

According to Ormalá, the biggest problem facing science in Finland is how to reorganize research as an integral part of the European system. "If you look at R&D in Finland, it is not international, it is very domestic, and in many fields it is of moderate quality," he says. "We have to learn that it is not sufficient to be good here; you have to be good at the international level."

Awareness of the value of international cooperation in stimulating the quality of domestic research has been one of the main reasons behind Finland's recent departure from its traditional neutrality and its increasing willingness to participate in international research projects. "If you are trying to get research money from Brussels, then you must be good," says Paavo Uronen, professor of physics and vice-rector of the Helsinki University of Technology. "Otherwise you will not have research groups which will survive under international competition. Competition will ensure that the quality of research is good enough."

Finland became a full member of CERN (the European Laboratory for Particle Physics), for example, in 1991. And it is applying for full membership of the European Space Agency, of which it is an associate member. "For a small country these programmes mean a lot; we have been very isolated in the past, and it will certainly change the climate", says Helander at the academy.

Both decisions have been actively promoted by politicians keen to see Finland integrated into Europe. But both have been controversial in the scientific community. In both cases, the subscriptions will come, *de facto*, from the domestic research budget; and there are those who argue that, in strictly economic terms, better value might be obtained from pursuing other goals.

So far, however, despite the fears of a shift away from investment in basic research and into joint European programmes, there has been little protest about the impending cuts from among the scientific community. "The general reaction has been calm," says academy president Tanskanen. "The economic situation is so bad that everything has to be compared to that". □

Universities tighten their belts

AFTER several decades of rapid breeding, the Finnish university system has so many children that it doesn't know what to do. As a result of intensive regionalization, the country now boasts 17 universities, 3 art schools and 22 newly designated polytechnics; that makes a total of more than 40 university-level institutions in a nation of only 5 million inhabitants.

In the good times, this was not a problem. As late as 1987, the government — buoyed by the apparent strength of the economy — committed itself to increasing spending on universities by 15 per cent in real terms for the following four years. This ambitious goal has almost been met. Government spending on universities rose from Fmk3.23 billion (£425 million) in 1987 to Fmk4.78 billion in 1991 in constant prices, an increase of almost 50 per cent. The money has allowed all universities to buy research equipment and fill new teaching positions and professorial chairs.

Now the day of reckoning has arrived. With the collapse of the Finnish economy, the government has higher education in its sights as it seeks to put a tight squeeze on public spending. It has already announced a 10 per cent cut in the amount of money the universities will receive next year, and 5 per cent cuts in each of the following two years. It is also planning a major restructuring of the university system, due to be announced next April, which could mean some of the smaller universities giving up a significant part of their research activities.

Part of the budget reductions will come from the reduction in public sector salaries. But reductions elsewhere will be much larger, resulting in major cuts in research budgets and running costs, and the deferment of capital spending.

The University of Helsinki, for example, founded 350 years ago and the largest in the country, is cutting its spending on scientific equipment by 27 per cent next year and its operating expenses, including journal subscriptions and laboratory materials, by 14 per cent.

To Risto Ihmuotila, professor of agricultural science and recently appointed rector of the university, the basic problem is straightforward. "We have too many universities," he says, and he has publicly proposed reducing the total by about half. But Ihmuotila, a former chairman of the National Council for Higher Education, admits that such a solution would be unacceptable in political circles. The alternative now gathering support among many university administrators is for greater selectivity and concentration of both teaching and research resources.

"The only way that our science and

technology will survive at an international level is to have centres of excellence," says Paavo Uronen, professor of physics and vice-rector of the Helsinki University of Technology. "Our system was expanded too fast and in too many places. Now each university wants to stay independent and keep their activities."

Uronen would like to see each university boasting its own strong field. Such a proposal is widely expected to be endorsed next year by the Ministry of Education. The idea is that universities should concentrate on groups that are particularly strong in research, and organize their finances in such a way that support for these groups is increased — where possible from industry — and that money for less productive groups should be decreased or eliminated.

Many, however, already face the problem that, with the current economic



Helsinki's rector Risto Ihmuotila bridges at the 'public sector' treatment

recession, there has been a significant decline in the amount of money that private industry is prepared to provide to university research groups. This is particularly difficult for the newly created universities of technology.

Reduced government spending is also concerning the Academy of Finland. "If universities cut down, then we have to cut down too," says Elisabeth Helander, the academy's research director. "The problem is that figures from the universities show that about half of the professors will retire in the five years either side of the end of the century. That means that it is important to keep young people in the system that you can recruit from; a lack of funding will make this difficult."

Ihmuotila at Helsinki University claims that the major problem now facing universities is that they are being treated as part of the public sector, and are being given identical treatment to other parts of this sector. "We think that universities are always needed, and that our task is even more important than it was earlier," he says. "Our economic life is based too much on the metal and forestry industries; in the future we will need more high-technology industries, based on electronics and the biosciences, and it is support for these type of areas that the universities can provide". □

Technology dodging the cuts

"THIS is a changed culture compared to 20 years ago," says Juhani Kuusi, director general of Finland's Technology Development Centre (TEKES). "Finland has had a very humanistic tradition and we are young in technology; but all that is changing very fast."

Official statistics confirm the change. The share of high-technology products in the nation's exports rose from 4 per cent in 1980 to 15 per cent last year. Much of the credit for this is claimed by the two agencies that have been responsible for boosting technical research during this period: TEKES, set up in 1983 to lead the nation's technology strategy, and the Technical Research Centre (VTT), the largest research centre of its type in the Nordic countries.

Both agencies bear witness to the government's past commitment to creating a national technology policy, but are now having to adapt to a harsher economic and political climate.

VTT is the older of the two. Established more than 50 years ago as a testing facility for the nearby Helsinki University of Technology, VTT is responsible for carrying out technical research, both on its own initiative and increasingly under contract to private industry, in areas ranging from the processing of forest products to molecular electronics. VTT grew rapidly in the 1970s and 1980s, with its research budget (and personnel) doubling in each of these two decades. It now has a budget of Fmk950 million (£130 million) financed partly by the government and partly by contract research raised from private industry. Its 2,700 staff members, 55 per cent of whom hold university degrees, now work in 34 separate laboratories; most of these are still based in Espoo, but others are distributed in major cities throughout the country.

The period of growth has now come to an end. Staff numbers are already being cut as a result of a government decision to reduce the number of public sector employees. Contract research funding from private industry is becoming scarcer as a result of the recession. And VTT expects a cut in its government grant for next year.

Despite this, VTT director general Markku Mannerkoski is determined to remain optimistic. "We need to adapt to a new situation," he says. "The recession serves to provide further incentive for streamlining our activities and making further efforts." In practice, that means that private industry is likely to be invited to take a more active role in determining the content of research programmes at VTT. "If industrial enterprises become interested in participating in strategic

research and technology programmes to advance generic technologies, that means more commitment from industry to these programmes."

TEKES faces an equally daunting task in adapting to the new climate. So far, it has avoided major budget cuts. While basic research funded through the Finnish Academy and the universities is being reduced in line with general reductions in public spending, TEKES has persuaded the government to give it a 10 per cent



TEKES director Juhani Kuusi: "The big challenge will be how to get back the money we will be paying into the Framework programme".

increase next year.

The decision reflects the important role that TEKES now plays in supporting research projects in private industry. When the agency was set up in 1983, it was given responsibility for 12 broad-based government-led programmes covering familiar generic technology areas such as biotechnology and information technology. On their completion,

however, many of these programmes were criticized for failing to produce results which were of immediate interest to Finnish industry.

The emphasis of TEKES's activities has now shifted to include more direct funding of R&D projects within industrial enterprises — two-thirds of last year's budget of Fmk862 million was spent in this way — and giving industry greater control of individual technology programmes.

The second reason why TEKES remains central to the government's strategy is the role it has played in stimulating and coordinating research projects between Finnish scientists and those in other European countries, for example through EUREKA and the various programmes of the European Commission such as ESPRIT and RACE.

In 1991, for example, Fmk145 million out of TEKES's total budget was spent supporting such projects. TEKES is also responsible for a worldwide network of industrial attachés in its foreign embassies. "Our achievement has been the internationalization of Finnish technology," says Kuusi.

This task will increase in importance if, as expected, Finland becomes a full member of the European Commission's Framework programme next year, meaning that it will no longer be able to subscribe to projects on an individual basis, but will have to compete for funding from a central pool. □

Opposition confounds nuclear policy

COULD Finland become the first country to abandon its nuclear power programme through a political accident? This may yet be the consequence of a surprise vote in Parliament last month, in which the deputies agreed that nuclear power should not play any role in the country's future energy strategy.

The energy strategy was not meant to go like that. The government's plans were that the key vote on the future of nuclear power in Finland would be based on the application it received last year from power companies to build the nation's fifth 1,000 MW nuclear power plant. Although the licensing decision is the responsibility of the cabinet — which has already delayed any decision for 18 months — under Finnish law any such decision has to be ratified by Parliament.

But the whole process has now been overtaken by events. When Parliament debated a separate report on energy strategy produced by a parliamentary committee, a series of amendments were presented. One of these, from an opposition member, suggested that the nuclear

option should be abandoned. And, on a free vote, the amendment was adopted.

Party officials were taken by surprise. With ten separate parties in the Parliament — five forming the coalition government and the five others in opposition — confusion on a free vote was perhaps inevitable.

The whole issue is now in abeyance while the government decides what to do next. The delay has been widely welcomed by the country's growing anti-nuclear movement, but the nuclear industry is dismayed.

"The situation is quite a mess at the moment, and we have decided to wait and see what happens," says Anti Rooskannin, a spokesman for Imatran Voima (IVO), the country's leading producer of electricity, which has already compiled a short-list of six potential sites for the fifth reactor. The company had been confidently expecting the go-ahead. "Now we would prefer to wait for a positive decision, even if it takes time, rather than face a rapid negative one?" says Rooskannin. □

Low-temperature gains

IF Finland is known for its low temperatures, it is also becoming known for its low-temperature research. One of the leading groups in the field, with a growing international reputation, is the Low Temperature Laboratory at the Helsinki Technical University, based in the Helsinki suburb of Espoo.

A special interest of the laboratory has been the application of low-temperature physics to the study of the brain. Physicists at the laboratory have been working since the mid-1960s on ways of measuring the very weak magnetic fields in the femtotesla range produced by the electric fields flowing in the human cerebral cortex. From the early days they have been doing this with the use of SQUIDS (superconducting quantum interference devices), using these weak magnetic fields to map out neural activity in the brain.

This method, known as magnetoencephalography, allows noninvasive localization of the activated brain areas with a special resolution of a few millimetres and temporal resolution, under favourable circumstances, of a millisecond. Such measurements provide detailed insight into the organization and functioning of the brain.

A number of experimental devices have been produced in the laboratory for these purposes, starting with the world's first multichannel instrument for magnetic field mapping, a four-channel gradiometer, built in 1983. After two intermediary versions, using seven and 24 channels respectively, the latest device to emerge from the laboratory is a 122-channel neuromagnetic measuring system, made up of 61 two-channel thin-film units evenly distributed over the whole of the head, known as Neuromag-122.

The advantage of the new instrument is that it allows a complete set of data to be collected from an experimental subject without repositioning the instrumentation allowing processes in different areas to be followed simultaneously.

A small company has been set up to develop Neuromag. One fifth of the shares are owned by members of the research team, 30 per cent by Instrumentarium Corporation, Finland's major producer of medical instruments, and the rest by the Finnish National Fund for Research and Development (SITRA).

Although currently operating from the laboratory itself, the company is soon to relocate to its own premises, from which it hopes to sell the new machine to

university research groups throughout the world.

Marketing the new machine will be done through Instrumentarium. "It is obvious that the Neuromag machine is the final answer for this type of equipment, but in a small country like Finland our main difficulty is marketing," says Olli Lounasmaa, professor of physics and director of the department. "We have had many enquiries about this instrument and have been asked for several quotations, although so far we are waiting for a firm order."



Channel vision: The Neuromag-122 in action.

Lounasmaa says that, unlike in some other countries, there is little opposition to university scientists becoming involved in commercial activities. "In some countries, people are worried about the possibility of conflicts of interest with those working on government funds. But in Finland, it is just the opposite. The government is very happy about supporting this kind of activity in universities which will lead to commercial products."

As with virtually all other research groups in Finland, Lounasmaa says that he expects a slight reduction in his research budget next year. But he hopes to be able to take advantage of two policy changes: one is a greater concentration of research funds in university-based centres of excellence, the other is the potential availability of funding from Brussels through the Human Capital fund.

Two applications from his group — one in biomagnetism and the other in ultra-low-temperature physics — have already received high marks from scientific reviewers, and both projects are candidates for funding early next year.

"It is always a difficult problem for a small country that, if it joins a large programme such as the European Space Agency, it finds that the big countries — Germany, France and the United Kingdom — tend to get the most interesting parts, and small countries are very likely to get the bits and pieces which no one else is interested in, with the danger of just becoming a footnote. The attraction of these new applications is that the installation will be in Finland, and we will have control over them." □

Talking to Europe

SEPARATED by vast distances and hostile terrain, it is perhaps not surprising that the scattered inhabitants of the Nordic countries have become the world's greatest enthusiasts for mobile telephones. A survey earlier this year showed Sweden, Finland and Norway topping the international league with between 6 and 7 per cent of the population owning such telephones in each country; in Britain and the US, the figure is just over 2 per cent, and in Germany less than one.

Having a strong market on its doorstep has been an important key to the success of the mobile telephone business of Finland's Nokia Corporation. And this itself has been made possible by agreement in the early 1980s between the four Scandinavian countries of a common mobile network through agreement on a single operating standard. Both experiences have encouraged the company to become involved in many of the high-technology research programmes, such as ESPRIT and RACE, organised by the European Commission in Brussels.

According to Viljo Hentinen, Nokia's vice-president of research and development, 40 per cent of the company's spending on R&D — which, at Fmk 933 million (£130 million) last year represents 6 per cent of the company's annual turnover, and 20 per cent of the total R&D spending by Finnish industry — is now done through international projects. This is partly to collaborate in pre-competitive research, and partly to enable the company to participate in the setting of international standards.

After a period of rapid growth in the 1980s, the company suffered a loss last year, but telecommunications and mobile telephones have remained two bright spots. Both represent the fruits of heavy investment in research and development. In the case of mobile telephones, where the company has pioneered the use of digital cellular telephone systems and is now the second largest producer in the world (after Motorola), R&D represents 10 per cent of the division's turnover.

Hentinen admits that Nokia does not spend much of its research money with university departments — at present, less than 2 per cent of its total R&D budget. But it is still keen to benefit from the fruits of Finnish research, and its Tampere and Espoo laboratories work closely with the nearby universities.

Hentinen is keen to see universities build on their top-quality research teams. "Universities which do high level research must create centres of excellence, hopefully in areas of interest to industry," he says. "This is the way of the future."

No biotechnology bandwagon

OVER the past ten years, stimulated by worldwide enthusiasm for biotechnology, Finland's scientific community has been laying the groundwork for a genetic engineering industry. But as these efforts bear fruit, the wind is being taken out of its sails by a virtual absence of risk capital and apparent lack of interest from industry in radical new production processes.

The situation has been a major source of frustration for scientists such as Helena Mäkelä, professor of infectious diseases at the National Public Health Institute. Mäkelä chaired a committee set up by the Finnish Academy in the mid-1980s which designed a five-year programme adopted by the Ministry of Education to boost biotechnology in Finland.

The academic part of the programme was highly successful. Almost 70 PhDs have been produced in the disciplines considered essential for the health of a successful biotechnology industry. Four university-based biotechnology centres have been created with special funds from the Ministry of Education, based on research at the universities of Helsinki, Oulu, Turku and Kuopio.

But industrial enthusiasm for biotechnology has been less forthcoming, and has cooled even further with the current recession. "In the academy, it has become clear that industry has not developed an interest in biotechnology as fast as we had expected, and that many of the scientists we have trained up will be left without a job", says Mäkelä.

Generous government funding for the university-based centres, combined with efforts to ensure that each is plugged into international research networks, has ensured that the standard of research and

training has been high. The Biocenter Oulu, for example, which was established in 1986 and brings together project groups based in a number of different research departments at the University of Oulu, already accounts for one third of the PhDs produced by the university.

At the University of Helsinki, more than a hundred people now work at the Institute of Biotechnology, set up in 1989 as a successor to the university's Recombinant DNA Laboratory, which was one of the first groups to carry out genetic engineering research in Finland. "Over the past ten years, a molecular biology culture, based on four centres each now enjoying an international reputation, has been developed in this country almost from scratch", says the institute's director, Mart Saar-ma.

The picture in terms of new products is not all gloomy. For example, scientists at the Biotechnical Laboratory of Finland's National Technology Centre have developed a technique for bleaching wood-pulp without the use of chlorine which is steadily gaining acceptance in the country's paper and pulp industry. Research carried out by Kari Kivirikko at the University of Oulu has led to the development of a new drug for fibrosis, clinical trials of which are now being carried out by Hoechst in Germany.

So far, however, such success stories have been the exception rather than the rule. The traditional conservatism of Finnish industry — particularly those, such as the paper and pulp industry and the brewing industry, in which the government had been hoping biotechnology would make its mark — has made it slow to accept the new techniques.

Obtaining the venture capital that entrepreneurs in other countries have used to build up small biotechnology companies has also proved difficult. "We have a very exciting research environment, but the problem in this country is a lack of venture capital", says Karl Tryggvason, professor of biochemistry at the University of Oulu and scientific director of the Biocenter.

On top of all this, there have been a number of setbacks that have increased the nervousness of potential investors. There was, for example, the experience of Genesit, a small biotechnology company set up in the mid-1980s with funding from the government, the Finnish Industrial Fund for Technical Development (SITRA) and seven large Finnish corporations to explore ways of exploiting the use of *Bacillus subtilis*. Genesit was widely seen as Finland's entry ticket into the brave new world of venture-capital-

based biotechnology. But it eventually folded after encountering technical difficulties in developing marketable processes.

Equally embarrassing has been the recent collapse of attempts of another biotechnology company, Pro Vivo, to establish a company, called Arctic Point, to produce food products based on blood extracts. Pro Vivo bought the technology from VTT, and persuaded a group of investors to put up Fmk25 million to build a new factory for producing sausages using globin extracted from blood as a substitute for milk protein.

According to Ralf Lundell, the managing director of Pro Novo, the project failed when it was discovered that the globin did not bind to the fats in the sausages as well as the conventional forms of protein. The companies responsible for marketing the sausages were not prepared to change the basic recipe to increase the amount of lean meat to compensate for this, he says. As a result, Arctic Point, widely touted as one of the successes of Finnish biotechnology, went into bankruptcy last month.

Now, even companies with an established track record in developing novel approaches to biotechnological processes are cutting back on long-term work to focus on more product-oriented objectives. Finland's state-owned brewery Alko, for example, which for the past 10 years has had a team of about 70 researchers working on the development of new vector systems such as *Trichoderma* for producing cellulase enzymes and other products, has decided that it wants its scientists to work more closely than in the past with its business units.

"We are focusing our work on the strategic interests of our business decisions, and our research goals are becoming more clearly defined", says Matti Korhola, director of research at Alko. In the process, the company will be reducing the size of its molecular biology research team by about 10 per cent. "We are maintaining a fairly strong centralized R&D function in the research laboratories, but we also aim at faster utilization and application of our results", says Korhola.

In the pharmaceutical field, several companies are turning away from genetic engineering, wary of the time needed to turn bright ideas into products, and refocusing on more conventional ways of producing new drugs.

Despite setbacks and retrenchment, however, many scientists remain optimistic about the future. "Over the past ten years, the number of people able to master the basic techniques of modern biotechnology have increased exponentially, and we now have the human capacity that was missing when we started," says Leevi Kääriäinen, research director at the Institute of Biotechnology in Helsinki. "The question now is how to keep the faith." □



The new Medipolis Centre at Oulu, housing start-up biotech companies.

By their teeth ye shall know them

An argument about a fossil suggests that the decade-old lessons of cladistics have not yet been fully appreciated.

If ever there was a good story in palaeontology, it is the one about the gradual evolution of mammals from reptiles. Some 280 million years ago, a reptile appeared that resembled mammals in certain key features: it was the earliest-known therapsid (M. Laurin & R. R. Reisz, *Nature* **345**, 249–250; 1990). Over the next 80 million years, characteristically mammalian features appeared among therapsids. In particular, the dentary (tooth-bearing) bone of the jaw began to predominate over the other jaw elements. Some of the later therapsids were astonishingly mammal-like in form.

In the Triassic Period, around 200 million years ago, the earliest 'true' mammals appeared. But the line between mammals and therapsids was by then so fine as to be barely discernible. The first mammals lived alongside the therapsids, which finally became extinct in the Jurassic around 160 million years ago.

Earlier this year, Fox *et al.* described the fragmentary dentary and teeth of what they claimed to be a therapsid (*Nature* **358**, 233–235; 1992). The presence of small, post-dentary bones is suggested by scars in the inner face of the bone. The teeth have elaborate cusps, but single, unbranched roots.

Fox *et al.* named this fossil *Chronoperates paradoxus*, an 'unexpected wanderer in time', because it lived in the Palaeocene, 100 million years after the last known therapsids were thought to have died out. Such a discovery was unexpected indeed.

Aware of the implications, Fox *et al.* discussed what, if anything, *Chronoperates* could be, if not a therapsid. The most likely alternative is a symmetrodont, a member of an extinct group of Mesozoic mammals. But although symmetrodonts have multi-cusped teeth, the roots of the lower cheek teeth have two roots apiece, not just one, and there are a number of other differences in detail. These differences, though, are 'subtle', as M. Novacek pointed out in the accompanying News and Views article (*Nature* **358**, 192; 1992).

H.-D. Sues went further, to dispute all the characters proposed by Fox *et al.* as indicative of therapsid affinity (*Nature* **359**, 278; 1992). Some therapsids have teeth with roots that are "incipiently divided", weakening the equation of single roots with the therapsid state, and

the scars for the post-dentary bones in *Chronoperates* are open to other interpretations (including post-depositional distortion.) Nevertheless, Sues stops short of suggesting what he thinks *Chronoperates* actually is, if not a therapsid — something seized on by Fox *et al.* in their reply on page 540 of this issue. The debate is likely to run and run, at least until more complete remains of this intriguing creature are unearthed.

But there is a deeper problem: that *Chronoperates* is a fossil out of time only if one assumes that the transformation series between reptiles and mammals happened in the way the textbooks would have us believe. It could be that the distinction between mammals and therapsids rests with the beholder: what with the capacity of the human mind to pick patterns from apparently random information, the transformation series between reptiles and mammals is just too good to miss.

Transformation series like this do not really work, because, unlike us, evolution has not the benefits of hindsight. In other words, transformation series 'create' evolutionary lineages after the fact. Consider: if the mammals (and *Chronoperates*) had vanished in the Palaeocene, would the transition between therapsids and mammals in the fossil record be so obvious? Would we not see, instead, a bush of competing, convergent lineages in which the features we now see as mammalian (thanks only to the abundance of modern mammals) appear more randomly? If so, then the line between therapsid and mammal, so long debated, would be seen as artefactual, and it would hardly matter whether *Chronoperates* is one thing or the other.

In the real world, of course, the mammals did evolve and diversify, but the very fact of their existence implies an ancestry in which the characteristic features of mammals appeared. It is no surprise, then, that these features can be recovered from fossils: it is easy to find such things, when one is convinced of the appearance of that for which one seeks.

Because of the recursion that crops up whenever transformation series are invoked, to so-called school of 'transformed cladists' suggested that the only objective way to reconstruct phylogeny is with information from extant species (D. Rosen *et al.*, *Bull. Am. Mus. nat. Hist.*

167, 163–275; 1981). This means that proposed transformation series can only be tested against phylogenies created with independent neontological or ontogenetic information. If fossils are to be included at all, they must be fitted in afterwards, as provisional groups called 'plesions' (C. Patterson & D. Rosen, *Bull. Am. Mus. nat. Hist.* **158**, 85–172; 1977). Above all, information from fossils cannot be used to overturn a phylogeny created using neontological data — not because they are unable to do so, but because such a course would be methodologically unsound. Based on these ideas, B. G. Gardiner proposed that of all the amniotes, mammals were most closely related to birds (*Zool. J. Linn. Soc.* **74**, 297–232; 1982). This idea was generally pilloried, and contested in print by a student of therapsids (T. S. Kemp, *Zool. J. Linn. Soc.* **92**, 67–104; 1988).

Moreover, J. Gauthier *et al.* used an extensive database to show that fossils could be used to overturn phylogenies based on neontological data, irrespective of whether it was right to do so (*Cladistics* **4**, 104–209, 1988; H. Gee, *Nature* **334**, 13–14; 1988.) The process turned a phylogeny of amniotes that owed much to Gardiner, into the more traditional scheme complete with the reptile-to-mammal transition. But when the analysis was repeated using taxa that were present only up to and including the Triassic — before the modern radiations of mammals — the transformation series was less apparent.

There is an illuminating literary parallel to all this. In his essay "Kafka and his Precursors", Jorge Luis Borges speculates on the diverse sources which could conceivably have influenced Kafka's own writing. "In each of these texts", he writes, "we find Kafka's idiosyncrasy to a greater or lesser degree, but if Kafka had never written a line, we would not perceive this quality; in other words, it would not exist The fact is that every writer creates his own precursors." (Original emphasis: translation by J. E. Irby).

It is Spike Milligan, though, who perhaps summarizes the debate about *Chronoperates* most succinctly, when recounting the reaction of tourists confronted by the eponymous hero of his children's story *The Bald Twit Lion*: "The tourists couldn't believe their eyes, some couldn't even believe their teeth". **Henry Gee**

Of mice and the MHC

Alan Grafen

Do mice recognize their kin? They are known to discriminate between conspecifics through the products encoded by the highly polymorphic loci of the major histocompatibility complex (MHC)¹; the MHC is not only involved in immune recognition, but also affects an individual's odour profile. There is, however, a general methodological reluctance to accept such discrimination as evidence of kin recognition². It was shown last year³ that female mice prefer mating with males that differ from them at the MHC locus and now, on page 581 of this issue⁴, the same authors (C. J. Manning, E. K. Wakeland and W. K. Potts) show that females nest communally with other females that are similar at the MHC. Smelling MHC differences seems to be important in the social lives of mice.

The technique of Manning and colleagues is to put the geneticist's knowledge of inbred strains, breeding designs and MHC-typing to the service of behavioural ecology. Seminatural conditions allow freedom of mating and communal nesting, and individual tags mean that the mice can be observed without disruption.

Around the time a female mouse gives birth, she usually finds a nestmate who has just given birth. The two females (sometimes more) suckle the pups indiscriminately. Manning *et al.* compared the nestmate that was chosen by each female with the other females that could have been chosen instead. They found that the MHC similarity was significantly greater with the chosen nestmate. The genetic manipulations allowed relatedness and MHC similarity to be uncorrelated in some of the groups, so the preference for MHC could not be explained by females choosing relatives according to a general background of genetic similarity. The MHC locus stood out as the focus of choice. These results should encourage workers on kin recognition in other species to investigate the importance of genetic discrimination in natural conditions.

We have one well-documented example of kin recognition in the ascidian *Botryllus schlosseri*, larvae of which tend to settle with histocompatible individuals⁵. Can mice be taken as a second example? Certainly Manning *et al.* have shown that the discrimination is used in nature, and this in itself is very important. It is still worth asking whether it counts as kin recognition, the essential question being whether mice choose nestmates according to MHC alleles in order to help relatives.

Manning *et al.* refer to "expectations

that kin will be preferred as nesting partners". In a recent theoretical study⁶, Giraldeau and Caraco have tackled the issue of whether kin selection should lead to kin groups, and the answer is subtle. There is no general presumption. Some concepts can be introduced with some starkly simplified examples based on the nesting of mice.

(1) Two sets of two sisters give birth on the same day, and nests are big enough for only two females. If success in rearing pups is unaffected by the



C. J. Manning

Suckling cousins? A single female mouse nurses the broods of her nestmates.

relatedness of nestmates, then there is no kin-selective reason why the sisters should pair up.

(2) Suppose there are two sets of two sisters and one extra female. There is a disadvantage to nesting alone, and two sisters should pair up if this makes it less likely that either of them will end up as the odd one out.

(3) Suppose there are two sisterhoods of three and four. A quick-counting member of the three would try to pair up with one of the four. When the others realized what was happening, and paired up with sisters, it would be one of the original four that was the odd one out.

These examples are meant to show that any warm cosy feeling animals may gain from grouping with kin is not enough. A kin-selective advantage must come about through differential effects on number of offspring of the options of pairing with kin and not doing so. They are also meant to show that in some circumstances it may actually be

advantageous to group with non-kin⁶.

Returning to the real world, should mice really prefer to nest with kin? It does seem likely that, in general, nesting with kin would reduce the chance that kin end up nesting alone. To this extent it would be churlish to deny mice parity with *B. schlosseri*. But this tentative acceptance as kin recognition would be placed in doubt if it were shown that helping kin was not the major selective force. For instance, a possible reason for preferring similarity at the MHC locus is that females may discriminate between their own pups using the MHC. A female may prefer to belong to a nest in which both nestmates find it more difficult to distinguish their own pups, as this may increase effective cooperation. That would be an interesting selective force, but not a kin-selective one.

A point not explicitly made by the authors can be drawn tentatively from their Table 2 (see page 582). A female seems to match the MHC of potential nestmates against a referent that is derived from the haplotypes present in her own natal nest, and not from her own genotype. More direct evidence of this would be welcome. For one thing, the workings of the recognition system are of interest in themselves. Further, if females choose in an evolutionary sense to use a natal nest referent rather than their own MHC alleles, this may reflect on the selective forces involved. Specifically, is natal-nest MHC similarity really the best cue females have for kinship?

One technical reservation concerns the breeding scheme used to derive the experimental mice in which relatedness and MHC similarity were uncorrelated. Only a few generations elapsed between the founders and the experimental animals. So far as I can see, it follows that a large stretch of chromosome on either side of the MHC locus is likely to share similarity with the MHC locus. Not enough crossovers have had time to occur to restrict the conclusions definitively to that locus rather than one nearby, but this point does not threaten the rejection of general genetic similarity across all chromosomes as the effective variable.

Recognition systems are harder to study in nature than in the laboratory. But they are much more informative too. □

Alan Grafen is in the Department of Plant Sciences, University of Oxford, South Parks Road, Oxford OX1 3RA, UK.

1. Singh, P. B., Brown, R. E. & Roser, B. J. *Nature* **327**, 161–164 (1987).
2. Grafen, A. *Anim. Behav.* **43**, 42–54 (1990).
3. Potts, W. K., Manning, C. J. & Wakeland, E. K. *Nature* **352**, 612–621 (1991).
4. Manning, C. J., Wakeland, E. K. & Potts, W. K. *Nature* **360**, 581–583 (1992).
5. Grosberg, R. K. & Quinn, J. F. *Nature* **322**, 456 (1986).
6. Giraldeau, L.-A. & Caraco, T. *Evol. Ecol.* (in the press).

Running rings round modellers

L. W. Esposito

THE extraordinary rings around Saturn are now known to have parallels around its neighbouring giant planets, each showing intricate structure and complexity which deepens their visual splendour. Astronomers at a recent gathering* grappled with the questions prompted by observations and modelling in an attempt to understand why planetary rings are the way they are. Neptune, visited by Voyager II in 1989, received the most attention, with more than half the papers devoted to its rings.

Until 1977, the only known ring system circled the planet Saturn. Since then, ground-based and spacecraft observations have revealed ring systems around each of the giant planets. The rings are composed of myriad small particles each on an individual orbit, with particles ranging in size from dust motes to stadium-sized boulders. Each particle orbits its planet in less than a day. The early models of rings as smooth, circular, symmetric systems have been shattered by the detailed observations: "All theoretical predictions have been wrong! Rings never behave as predicted by models" (A. Brahic, University of Paris). An especially difficult case has been provided by Neptune's most substantial ring, the Adams ring, which includes four longitudinally incomplete arcs superimposed on an almost invisible complete ring (Fig. 2). The combination of ground and Voyager observations shows that Neptune's arcs are stable over at least five

years, but current models cannot explain this stability against keplerian shear (the fact that the orbital speed decreases with distance from the planet: the ring particles slightly more distant should lap those inside to completely fill the circumference). Other serious problems include the perturbing effects of the so-called 'outer Lindblad' resonance with the orbit of Neptune's moon Galatea, and the effects of particle collisions (D. W. Foryta, C. Ferrari, Observatoire de Paris). The model proposed by Goldreich *et al.*¹ (where co-rotation resonances excited by the gravitational perturbations from the inclined orbit of Galatea confine the ring particles in longitude) is consistent with the Voyager observations, as shown by Porco², but does not explain the observed persistence. Brahic asks further, why are planetary arcs around the different planets so similar and what is their origin?

In July 1992, Uranus passed in front of a dim star denoted U102. Observations of this occultation (R. G. French, Wellesley College, Massachusetts) revealed the λ ring of Uranus (photographed by Voyager, but never before observed from Earth) on one side of Uranus, but not on the other. This detection is reminiscent of the first reports of the Neptune rings — it confirms the Voyager pictures which showed it to be another clumpy, dusty ring. Theoretical models of the archetypal clumpy ring, Saturn's braided F ring, show that the braids, clumps and kinks can be understood as results of gravitational perturbations from the nearby moons and that the moon Prometheus actually collides with the ring every 17 years (S. Giuliatti-Winter, Queen Mary and Westfield College, London). The structure seen in Saturn's Encke gap ringlet is clearly related to the moon Pan (F. Spahn, University of Potsdam). Wakes created by the near passage of Pan travel one or more times around the circumference of the ring and eight wakes are visible in the Voyager ring occultations (L. J. Horn, Jet Propulsion Laboratory, Pasadena). These models and observations show clearly that the moons excite the heterogeneity in nearby rings. Whether one should

never believe data without a good theory, as suggested by Goldreich (or exactly the opposite, as suggested by French), the study of planetary rings is in a much happier state: after initial confusion following the Voyager results, we now see a convergence towards a common picture.

This picture involves a close relationship between the rings and the many small moons orbiting nearby or among them. Both the structure and the origin of rings can be explained by a multitude of small moons, still unseen. Some philosophical support for these undetected moons comes from the recent discovery of the moon Pan in Saturn's Encke gap³, whose existence had earlier been pre-

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

FIG. 2 Voyager 2 view of Neptune's rings showing 'arcs' in the outer (Adams) ring that were probably created from a shattered former moon and held in place by unseen moons.

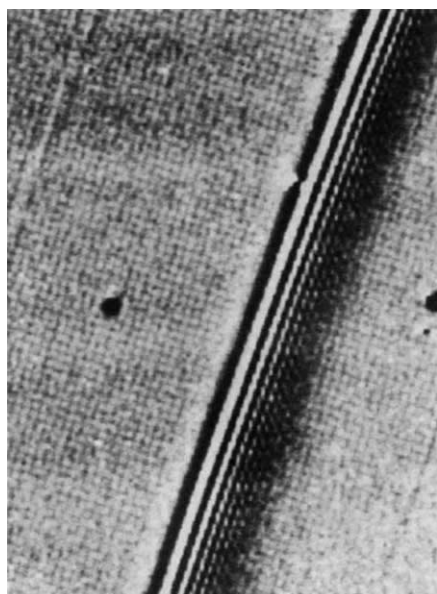


FIG. 1 Corrugations in Saturn's A ring raised by gravitational perturbations from Mimas propagate as a spiral bending wave, in a striking example of how moons sculpt rings.

dicted by its observed perturbations of the nearby ring. These moons are the natural result of a collisional cascade in which cometary impacts break the initial complement of planetary satellites into smaller objects, subsequently shattered themselves, and so on. This mechanism is common in all the ring systems, but because of its stochastic nature, the outcome is variable. Differences in the size of the destroyed moon and energy impact give different size distributions and amounts of material.

This process explains some of the differences and similarities among the known ring systems. The smallest particles become the rings; moons with radius 1–10 km perturb and confine the rings and provide reservoirs for dust. Calculations show that the present moons, rings and dust surrounding Neptune could be derived from such a cascade (J. Colwell, University of Colorado), and in fact hundreds of unseen small moons 1–10 km in size are predicted. Further, I suspect that Neptune's arcs were probably created in a single event, which destroyed a small moon and filled several co-rotation resonances near that moon's original longitude. The largest fragments help confine the arcs, as in Sicardy and Lissauer's model⁴.

* 24th Annual Meeting of the Division for Planetary Sciences, American Astronomical Society, Munich, 12–16 October 1992.

French called for more observations and deeper analysis of the Voyager data. Responses followed in the rest of the session: Saturn's icy rings have been polluted by infalling meteorite dust (J. Cuzzi, NASA Ames); the particles in Uranus's rings are larger than indicated by our best theory for their uniform precession (E. Marouf, Stanford University); the size and number of dust particles in Neptune's rings varies from point to point (M. Showalter, Stanford). D. P. Hamilton and J. Burns (Cornell University) showed that Saturn's dusty E ring regenerates itself: micrometre-sized particles from Saturn's moon Enceladus are accelerated by solar radiation pressure, recollide with Enceladus, and release a spray of new particles. A persistent ring is the result. The first observations of a stellar occultation by rings with the Hubble Space telescope refines our determination of the direction of Saturn's pole (M. Cooke and A. Bosh, Massachusetts Institute of Technology).

P. Nicholson (Cornell) described future occultations to be observed by the Visual and Infrared Mapping Spectrometer aboard the NASA-ESA Cassini mission, to be launched to Saturn in 1997.

An emerging picture of rings as youthful, stochastic structures sculpted by nearby small moons was shared by many of the participants. Open questions include the details of these interactions and the combined history of rings and moons suffering bombardment by interplanetary meteoroids. As data and theory continue to challenge each other, we can expect exciting new results. □

Larry W. Esposito is at the Laboratory for Atmospheric and Space Physics, University of Colorado, Boulder, Colorado 80309, USA.

1. Goldreich, P. *et al. Astr. J.* **92**, 195–198 (1986).
2. Porco, C. C. *Science* **253**, 995–1001 (1991).
3. Showalter, M. R. *Nature* **351**, 709–713 (1991).
4. Sicardy, B. & Lissauer, J. J. *Adv. Space Res.* **12**(11), 97–111 (1992).

NITROGEN FIXATION

New light on nitrogenase

R. N. F. Thorneley

PHOTOSYNTHESIS and nitrogen fixation are the two basic life processes that feed mankind. Understanding of photosynthesis was greatly increased by the Nobel prize-winning structure of the reaction centre of the purple bacterium *Rhodospseudomonas viridis*^{1,2}. Now, on page 553 of this issue³, Kim and Rees unveil a crystallographic structure at 2.7-Å resolution of the nitrogenase molybdenum-iron protein from *Azotobacter vinelandii*, identifying five similarities between its structural organization and that of the photosynthetic reaction centre (two types of subunit; unidirectional electron transfers; buried reducible substrate site; separate proton and electron transfer pathways; and inefficient energy coupling in light harvesting and ATP hydrolysis). They also propose a model for the 'docking' of the MoFe-protein with its electron-transfer and energy-transducing (ATP hydrolysis) partner, the nitrogenase Fe-protein; Rees's group has also recently solved the structure of that protein (at 2.9-Å resolution)⁴.

Biological nitrogen fixation, the conversion of dinitrogen (N₂) to ammonia, is a reaction that is currently restricted to certain prokaryotic organisms (diazotrophs) able to synthesize the enzyme nitrogenase. It requires a high energy input with at least 16 equivalents of ATP being hydrolysed for each N₂ fixed. The energy (ATP) comes either directly from photosynthesis occurring in the same organism (for example, cyanobacteria

and other photosynthetic bacteria) or indirectly from reduced carbon substrates supplied, in the case of the commercially important bacteria of the genus *Rhizobium*, by the legume plants (soya bean for instance) on which they form beneficial root nodules. Increased biological nitrogen fixation will increase food production, particularly in the Third World, while decreasing costs and environmental pollution caused by fertilizer run-off into surface and ground water.

The nitrogenase MoFe-protein is an $\alpha_2\beta_2$ tetramer of relative molecular mass about 240K, the α - and β -subunits being composed of 491 and 522 amino acids respectively. The tetramer contains 30 Fe and 2 Mo atoms, which are present as two types of metal cluster whose structures have been modelled by Kim and Rees⁵. Each $\alpha\beta$ -pair of subunits contains a 'P-cluster', comprising two 4Fe–4S cubes which are held together by two cysteinyl sulphur bridges between Fe atoms and possibly a third disulphide bond between two of the sulphides at adjacent corners of each cube. Each subunit contributes not only a bridging cysteine but also two other cysteines which, by coordinating to four Fe atoms located at the corners of the cluster, hold it at the $\alpha\beta$ -subunit interface.

These clusters are unusual, compared with Fe–S clusters present in other proteins, in that Mössbauer^{6,7} and electron paramagnetic resonance (EPR)⁸ spec-

tros copy have shown that in the resting enzyme all the Fe is in the ferrous state. Attempts to prepare stable model complexes containing (4Fe–4S)⁰ clusters have so far failed, presumably because of the high negative charge density. In nitrogenase, Kim and Rees show that, although the P-clusters are buried in a predominantly hydrophobic environment, each pair is located at the amino-terminal ends of six helices which could provide the necessary electrostatic stabilization. Transient kinetic and EPR data⁹ suggest that the P-clusters become oxidized by electron transfer to the second type of metal centre, the FeMo-cofactor, at the crucial point in the catalytic cycle when dinitrogen (N₂) is irreversibly committed to be reduced (fixed) to ammonia.

The FeMo-cofactor is probably the site of N₂ binding, although it is still not clear whether the N₂ interacts primarily with the Mo, Fe or S²⁻ components. The chemical studies of Chatt and others¹⁰ suggest that N₂ activation occurs at the Mo, but the structure of the FeMo-cofactor⁵ points to the possibility of activation at Fe. Six of the seven Fe atoms in the cofactor have trigonal pyramidal coordination geometry and are therefore coordinatively unsaturated and potential sites for N₂ activation. The recent activation of N₂ to yield NH₃ on a Fe(0) complex¹¹ provides further support for this notion.

The FeMo-cofactors are located in the α -subunits at the boundary of three domains reminiscent of the 4Fe–4S site in aconitase. Assuming that the Fe-protein docks to the MoFe-protein as modelled by Kim and Rees³, it seems likely that an electron transfer pathway runs from the Fe-protein initially to a P-cluster and then to its FeMo-cofactor partner. The distances involved, 10–14 Å for each hop, and the rate of oxidation of the Fe-protein (200 s⁻¹), are consistent with theories of electron transfer through protein matrices¹².

The key to understanding how nitro-

1. Deisenhofer, J. & Michel, H. *Angew. Chem. Int. Ed. Engl.* **28**, 829–847 (1989).
2. Huber, H. *Angew. Chem. Int. Ed. Engl.* **28**, 848–869 (1989).
3. Kim, J. & Rees, D. C. *Nature* **360**, 553–560 (1992).
4. Georgiadis, M. M. *et al. Science* **257**, 1653–1659 (1992).
5. Kim, J. & Rees, D. C. *Science* **257**, 1677–1682 (1992).
6. Smith, B. E. & Lang, G. *Biochem. J.* **137**, 169–180 (1974).
7. Zimmermann, R. *et al. Biochim. biophys. Acta* **537**, 185–207 (1978).
8. Smith, B. E., Lowe, D. J., Chen, G.-X., O'Donnell, M. J. & Hawkes, T. R. *Biochem. J.* **209**, 207–213 (1983).
9. Thorneley, R. N. F., Ashby, G. A., Fisher, K. & Lowe, D. J. in *Molybdenum Enzymes, Cofactors and Models* (eds Stiefel, E., Coucouvanis, D. & Newton, W. E.) (American Chemical Society, in the press).
10. Chatt, J., Pearman, A. J. & Richards, R. L. *J. chem. Soc. Dalton Trans.* 1852–1860 (1977).
11. Jimenez-Tenorio, M. & Leigh, G. J. *J. Am. chem. Soc.* **113**, 5862–5863 (1991).
12. Moser, C. C., Keske, J. M., Warncke, K., Farid, R. S. & Dutton, P. L. *Nature* **355**, 786–802 (1992).
13. Thorneley, R. N. F. *Phil. Trans. R. Soc. B336*, 73–82 (1992).

genase functions is the mechanism of coupling of Mg-ATP hydrolysis to proton and electron transfers. Free access of protons to the low-potential metal site(s) required to bind and activate N_2 is likely to result in uncontrolled H_2 production; yet protonation of N_2 is essential for NH_3 formation. With the MoFe-protein structure³ we can begin to address this problem, for it defines the possible access routes for N_2 , electrons and protons to the FeMo-cofactor.

The similarities, both structural⁴ and

kinetic¹³, between nitrogenase and at least two other ATP/GTP-dependent energy-transducing systems, the oncogene product p21^{ras} and actomyosin, are intriguing. They give studies on nitrogenase structure and function a significance well beyond that of nitrogen fixation. □

R. N. F. Thorneley is at the AFRC Institute of Plant Science Research, Nitrogen Fixation Laboratory, University of Sussex, Brighton BN1 9RQ, UK.

GLOBAL CHANGE

Lessons from past climates

Eric J. Barron

THE prediction of future climate change is based largely on general circulation models (GCMs) developed within the constraints of modern observations. The short-term nature of these observations, on top of natural uncertainties about the processes involved, give ample reason to question predicted responses to large perturbations, such as a doubling of atmospheric carbon dioxide. But on page 573 of this issue¹, Hoffert and Covey use data from two very different periods of the Earth's history to show that the estimated climate sensitivities are not unreasonable.

The Intergovernmental Panel on Climate Change (IPCC) has concluded that a global warming of between 1.5 and 4.5 °C is likely for a doubling of carbon dioxide. The majority of GCM projections fall within this range. Much of the range in estimates is attributed to uncertainties associated with cloud-climate feedbacks. The primary objective of Hoffert and Covey is to use two case studies from the Earth's history to derive an estimate, independent of GCMs, of the climate's sensitivity to a change in atmospheric carbon dioxide concentration. The authors consider two extreme climates, the Last Glacial Maximum 18,000 years ago and the mid-Cretaceous warm maximum (around 100 million years ago), each of which is characterized by differences in carbon dioxide relative to today.

Two types of information are essential in exploiting palaeoclimatic data in this way. First, the temperature of the Earth in each of the case studies must be reasonably established. At a minimum, the globally averaged surface temperature must be defined. But the seasonal and latitudinal distribution of temperatures would provide a firmer basis for interpretation of the results. Second, all the forcing factors (such as the concentration of greenhouse gases, changes in the Sun, aerosol content of the atmos-

phere, geography and vegetation) must be known. Lack of knowledge of any one significant climatic forcing factor would bias the estimate of the climate sensitivity to changes in carbon dioxide.

Obviously, estimation of any of these quantities for the past is associated with substantial uncertainty. For example, estimates of the globally averaged surface temperatures at the Last Glacial Maximum range from 3 to 5° C below the present-day average (15 °C); and mid-Cretaceous temperature estimates range from 6 to 12 °C higher. The distribution of polar ice at the Last Glacial Maximum, significant in defining the global albedo of the time, is still a matter of debate. The level of atmospheric carbon dioxide during the mid-Cretaceous is estimated to be anything from 2 to 11 times the present-day concentration. A value for the solar luminosity during the Cretaceous can be based on the evolution of that type of star, but this approach is basically unverified. In summary, neither the forcing factors nor the globally averaged surface temperatures are precisely defined for the past.

Hoffert and Covey confront these difficulties by using 'best' estimates of the temperatures and the forcing factors and then incorporating component uncertainties. From the best estimates, the Last Glacial Maximum and mid-Cretaceous yield climate sensitivities of 2.0 ± 0.5 and 2.2 ± 0.7 °C, respectively. Colder estimates of the glacial temperatures (5 °C lower than the present) would yield a sensitivity closer to 3.0 °C. But in short, the two eras yield sensitivities that are within the range cited by the IPCC. Certainly, the case studies suggest that the very low sensitivities suggested by some authors (such as Lindzen²) are not justified. But although the conclusions will be valuable in predicting the impact of fossil-fuel consumption and of the measured increase in atmospheric car-

bon dioxide, the precision of Hoffert and Covey's estimates should be treated with caution.

First, the authors' estimates of climate sensitivity fall within the lower half of the values generated by GCMs and cited in the IPCC report. Yet, so far not a single palaeoclimate simulation for the past has over-estimated the warming or cooling reconstructed from the geological data. In fact, palaeoclimatic simulations are notable in that they consistently underestimate the data, so that Hoffert and Covey may be underestimating climate sensitivity. The propensity of models to underestimate the data has prompted authors to speculate on the need to consider additional forcing factors, to re-evaluate the observations or to consider missing or inadequate feedback relationships in model construction.

The explanation of this enigma may be that GCMs have rarely been applied to the past; simulations yet to be tried may give overestimates. But two other explanations are possible. Hoffert and Covey's estimates may be correct if the surface albedo changes (ice cover and so on) are specified for the past, but if these feedbacks are calculated as part of the climate response rather than specified, then the lack of model sensitivity becomes apparent. Alternatively, the inadequacies of estimates of model sensitivity derived from global temperatures may become apparent only when the comparisons of data and models are based on seasonal and latitudinal data, not averages.

Second, Hoffert and Covey consider only two case studies. At least two other periods, the Younger Dryas³, at the end of the last ice age, and the Eocene⁴, are thought to be characterized by very different climate roles for the deep ocean. The mid-Cretaceous may also be characterized by a reversal of the deep circulation in response to increased atmospheric carbon dioxide⁵. But GCMs used for the IPCC report do not incorporate any explicit role for the oceans.

Hoffert and Covey's study adds to the weight of evidence in support of the IPCC conclusions. However, until GCMs actually simulate past climates, or, better, overestimate past changes in climate, we cannot be completely certain that GCMs incorporate the full range of climate sensitivities recorded in the geological record. □

Eric J. Barron is in the Earth System Science Center, Pennsylvania State University, University Park, Pennsylvania 16802, USA.

1. Hoffert, M. I. & Covey, C. *Nature* **360**, 573–560 (1992).
2. Lindzen, R. S. *Bull. Am. Met. Soc.* **71**, 288–299 (1990).
3. Broecker, W. S. *et al. Nature* **315**, 21–26 (1985).
4. Barron, E. J. *Paleoceanography* **2**, 729–739 (1987).
5. Barron, E. J. & Peterson, W. H. *Paleoceanography* **5**, 319–337 (1990).

A signal chain of events

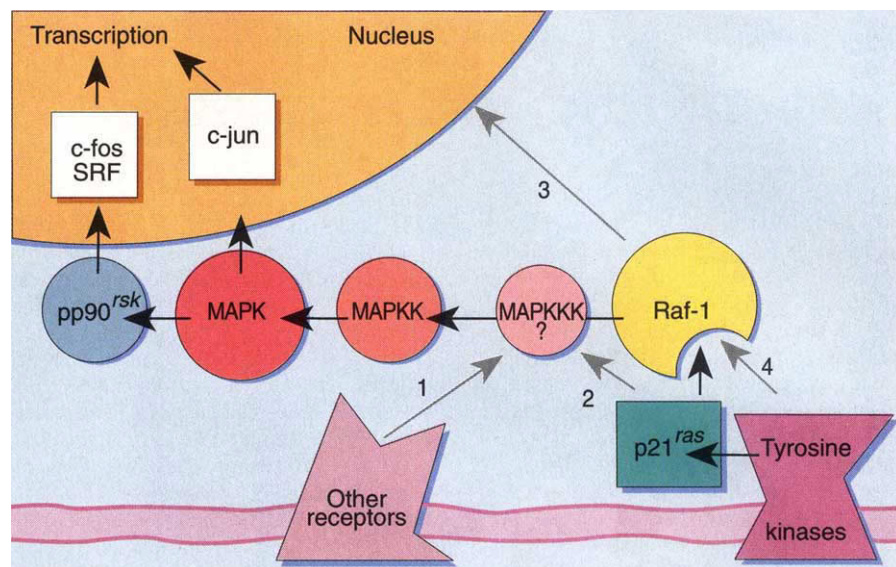
Thomas M. Roberts

It has been apparent for some years that molecular information arriving at the surface of the cell is transmitted as a signal from activated tyrosine kinases at the cell surface through serine/threonine kinases in the cytoplasm to reach (among other targets) transcription factors in the nucleus. The paper by Dickson *et al.*¹ on page 600 of this issue, among other recent reports, allows us at last to visualize details of one key pathway: the flow of signals via the proto-oncogene $p21^{ras}$ to Raf-1 kinase, thence to MAP (Mitogen Activated Protein) kinases and $pp90^{rsk}$ (ribosomal S6 kinase), and finally on to nuclear proto-oncogene products. As shown with black arrows in the figure, this pathway links many proto-oncogene products and is likely to form the main route for signal transduction from cell-surface tyrosine kinases. Although such simplicity is attractive, real life is undoubtedly more complicated.

The positioning of Raf-1 was central to understanding this pathway and five lines of evidence place Raf-1 downstream from $p21^{ras}$ in this flow of signals. First, *v-ras* was shown to mimic mitogen receptor activation and the actions of *v-src* in causing the hyperphosphorylation of Raf-1 (ref. 2). Subsequently, Kolch *et al.*³ showed that disrupting the function of Raf-1 could block the transforming activity of *v-ras*. Recently, Wood *et al.*⁴ showed the converse, that inactivation of *ras* could block the ability of a variety of receptors to affect Raf-1. Williams *et al.*⁵ showed, in biochemical studies, that maximal activation of Raf-1 autokinase activity required the activity of $p21^{ras}$. The final argument for arranging tyrosine kinases, $p21^{ras}$ and Raf in a linear pathway has just emerged from analysis of the *Drosophila sevenless* and *torso* receptor tyrosine kinase systems¹. In the *torso* system it was clear that Raf-1 was essential for signalling⁶ and in the *sevenless* system it was clear that $p21^{ras}$ was similarly essential⁷. Now Dickson *et al.* and N. Perrimon and colleagues (unpublished data) have shown that $p21^{ras}$ lies in the signal pathway to Raf-1 in both systems.

Positioning MAP kinases (MAPK) in the signalling chain resulted from two sets of experiments. $p21^{ras}$ is essential for the activation of MAPK by a variety of receptors, as shown in experiments using a dominant interfering allele of *ras* to block MAPK activation by mitogens and growth factors, and by the ability of *v-ras* to activate MAPK in the absence of mitogens^{4,8,9}. This finding raised the possibility that, as both MAPK and Raf-

1 were regulated by $p21^{ras}$, one of the two kinases might regulate the other. Three groups have now shown that Raf-1 can indeed activate MAPK¹⁰⁻¹². MAPK is activated in cells expressing *v-raf*, although the effect is not direct. Raf-1 cannot phosphorylate MAPK, but active Raf kinases can turn on prepara-



A key signalling pathway from tyrosine kinases at the cell surface to transcription factors in the nucleus.

tions of an activator kinase which is upstream of MAPK. This activator is termed MAPKK (MAP kinase kinase); hence Raf-1 might have a role as a MAPKKK (MAP kinase kinase kinase). But as none of the experiments used purified MAPKK, it is possible that at least one other intermediate is involved (hence the uncertainty over MAPKKK in the figure). The final links in this part of the chain come from experiments showing that $pp90^{rsk}$ is phosphorylated and activated by MAPK^{4,13,14}, and evidence that both kinases can translocate to the nucleus and can, at least *in vitro*, phosphorylate transcription factors such as Fos and Jun¹⁵⁻¹⁷.

There are, however, a number of experiments that fail to fit this nice linear arrangement (grey arrows in the figure). First, there are cell lines in which expression of *v-ras* fails to activate MAPK¹⁸ and ligands that activate MAPK even in the presence of dominant negative *ras*. This suggests that *ras*-independent means of activating MAPK exist (arrow 1). There are also cell types such as PC12 where *v-ras* activates MAPK but *v-raf* does not⁴. This suggests that there are *raf*-independent pathways from *ras* to MAPK (arrow 2); and, as *v-raf* activates immediate early genes in PC12 cells, there may also be a MAPK-independent route from Raf-1 to the

nucleus (arrow 3). Although these results may simply represent variations in the wiring schemes in different cells, it is possible that the apparently linear pathway is really a network. Indeed, a branched network might be expected from functional considerations, as it would allow integration of multiple cell-surface signals, which would occur most often *in vivo*.

A final question is: why might a branched network appear linear under certain experimental conditions? The study

of Williams *et al.*⁵ on the activation of Raf by *src* and *ras* may provide a clue. They found that Raf-1 was maximally activated by two synergistic signals, one *ras*-dependent and one *ras*-independent (arrow 4). Either signal alone activated Raf-1 at 10% of the level seen using the two together. In such a synergistic system, one of the signals can seem all important. Thus, blocking one of the activating signals knocks out 90% of the

- Dickson, B., Sprenger, F., Morrison, D. & Hafen, E. *Nature* **360**, 600-603 (1992).
- Morrison, D. K., Kaplan, D. R., Rapp, U. & Roberts, T. M. *Proc. natn. Acad. Sci. U.S.A.* **85**, 8855-8859 (1988).
- Kolch, W., Heidecker, G., Lloyd, P. & Rapp, U. R. *Nature* **349**, 426-428 (1991).
- Wood, K. W., Sarnecki, C., Roberts, T. M. & Blenis, J. *Cell* **68**, 1041-1050 (1992).
- Williams, N. G., Roberts, T. M. & Li, P. *Proc. natn. Acad. Sci. U.S.A.* **89**, 2922-2926 (1992).
- Ambrosio, L., Mahowald, A. P. & Perrimon, N. *Nature* **342**, 288-291 (1989).
- Simon, M. A. *et al. Cell* **67**, 701-716 (1991).
- Thomas, S. M. *et al. Cell* **68**, 1031-1040 (1992).
- Robbins, D. J. *et al. Proc. natn. Acad. Sci. U.S.A.* **89**, 6924-6928 (1992).
- Kyriakis, J. M. *et al. Nature* **358**, 417-421 (1992).
- Dent, P. *et al. Science* **257**, 1404-1407 (1992).
- Howe, L. R. *et al. Cell* **71**, 335-342 (1992).
- Chung, J., Pelech, S. L. & Blenis, J. *Proc. natn. Acad. Sci. U.S.A.* **88**, 4981-4985 (1991).
- Sturgill, T. W., Ray, L. B., Erikson, E. & Maller, J. L. *Nature* **334**, 715-718 (1988).
- Chen, R.-H., Sarnecki, C. & Blenis, J. *Mol. cell. Biol.* **12**, 915-927 (1992).
- Alvarez, E. *et al. J. Biol. Chem.* **266**, 15277-15285 (1991).
- Pulverer, B. J., Kyriakis, J. M., Avruch, J., Nicolakaki, E. & Woodgett, J. R. *Nature* **353**, 670-674 (1991).
- Gallego, C. *et al. Proc. natn. Acad. Sci. U.S.A.* **89**, 7355-7359 (1992).

activation — as is seen using dominant-negative *ras* in mammalian cells^{4,8,9}. Similarly, overactivation of one of the two input pathways with an activated oncogene can lead to (at least) partial activation of the downstream elements, as exemplified by the effect of *v-ras* on MAPK³. Variation in (for instance) the efficacy of *v-ras* in activating MAPK in different cell types might reflect changes

in the relative importance of input signals leading to choices between branches of a complex network rather than fundamental changes in the basic wiring of the network itself. □

Thomas M. Roberts is at the Dana-Farber Cancer Institute, 44 Binney Street, Harvard Medical School, Boston, Massachusetts 02115, USA.

POLYMER SCIENCE

Tender morsels for microbes

Paul Calvert

POLY[(R)-3-HYDROXYBUTYRATE] is a microbial storage polyester which forms in intracellular granules about 0.5 micrometres in diameter. The polymer has been commercialized as a biodegradable fibre-forming material with properties similar to those of polypropylene. Curiously, *in vivo* the polymer is soft and readily degradable, but after extraction it becomes a tough, stiff fibre. Why this change occurs has led to a disagreement. Sanders and co-workers, at Cambridge, suggest^{1,2} that some clever biological plasticization occurs *in vivo*, but de Koning and Lemstra argue³ that the small particles found in microbes are simply too small to crystallize and remain rubbery even when 150 °C below the melting point.

To put the difference in perspective it is worth considering the plant polymers cellulose and starch, which are both based on glucose. Cellulose, whose function is structural, has a regular arrangement with β -linked glucose units, which allow the polymer chains to be densely packed and highly crystalline. The impenetrability of the structure makes chemical and enzymatic digestion very slow. The α -linkage in starch, the energy-storage version of the polymer, drives the chain into an open helix, and frequent chain branching in the amylopectin fraction prevents extensive crystallization. The resulting material is soluble in water and readily degradable. Against this background, one would hardly expect poly[(R)-3-hydroxybutyrate] (PHB) to be an unbranched and readily crystallizable polymer, if it is to serve its biological function. But its behaviour when extracted shows otherwise. *In vivo*, some mechanism must operate to prevent crystallization.

Sanders and colleagues have shown¹ that PHB is an amorphous rubber in its native state within cells of *Alcaligenes eutrophus*. Carbon-13 NMR spectra show narrow lines, and measurements of spin-lattice relaxation times, (T_1), correspond to those of a rubbery polymer with a glass transition temperature of

about -40 °C. X-ray diffraction shows a completely amorphous polymer which becomes crystalline if the cells are washed with acetone and dried at 100 °C. A further paper² shows that mobility is partly lost on freeze-drying, although the polymer remains amorphous, and mobility is regained on rehydration. Careful rupturing of cells has no effect, but centrifugation of the granules causes them to crystallize. On this basis the group argues for some limited plasticization by water, which explains the changes on freeze-drying. In addition, the existence of an unknown hydrophobic nucleation inhibitor is postulated, either at the surface or distributed in the granule.

Lauzier *et al.* find⁴ that isolation of the granules results in a crystalline 'skin' on each granule but leaves a soft amorphous core. They attribute the persistence of the core to the presence of triacetin and other lipids, which could prevent further growth of the skin. But these lipids make up only about 0.5 weight per cent of the granule, which is nowhere near enough to solubilize a significant amount of the polymer. On the other hand, Lauzier *et al.* also find that the core stays soft down to liquid-nitrogen temperatures, which suggests that it is very plasticized.

de Koning and Lemstra³ have carried out scanning calorimetry on granules and freeze-dried cells. The glass transition of wet granules occurs at about -5 °C and moves up by about 5 °C on drying. This suggests that plasticization due to water is only modest. The difference in glass transition temperature reported by Sander's group and de Koning and Lemstra is not serious, as the NMR estimate involves a long extrapolation. It does, however, contradict the plasticized core deduced by Lauzier *et al.* Other evidence against the granule being heavily plasticized *in vivo* comes from Ceccorulli *et al.*⁵, who find that PHB can be plasticized with dibutylphthalate with a large reduction in the glass transition, but that there is a consequent increase in the

crystallization rate at room temperature.

It is the rate of nucleation that de Koning and Lemstra hold responsible for the pliability of the biological form of PHB. They estimate from nucleation-rate data that the typical time for a granule 0.5 μm across (like a natural granule) to nucleate for crystallization is 10^{12} seconds — over 30,000 years — at 20 °C. Compaction by centrifugation, as done in Sander's laboratory, welds the granules together into a single volume 10^{12} times larger, with a commensurate decrease in the time to nucleation. Once nucleated, the whole particle can be expected to crystallize. Indeed, this has been a problem in commercial forms of PHB, as the crystallized polymer tends to comprise a small number of large crystalline spherulites, owing to the low nucleation rate, and is therefore opaque and brittle. An extensive trial-and-error search has identified saccharin as a good nucleating agent (possibly because of a close match of intermolecular spacings) which improves the quality of the polymer^{6,7}.

Lastly, the granules *in vivo* are surrounded by some kind of coating, or possibly a lipid membrane. Treatments that damage the granule surface also cause crystallization. de Koning and Lemstra attribute this effect to adventitious particles which then contact the polymer surface. The earlier search for nucleating agents suggests that suitable particles are quite rare. The alternative explanation is that nucleation can occur at a water-polymer interface, whereas the membrane surface is non-nucleating, if not inhibitory. We know very little about the nucleating effects of different liquid-liquid surfaces, so anything is possible.

There has been little work on the crystallization behaviour of very small particles of synthetic polymers. Surface effects must be very important and there may be strange plasticization effects arising from the small size. The bacteria do seem to have hit upon a simple method for keeping polymers rubbery (and edible) at far below the melting point. If we could do this too, it would be a terrific way of making tough coatings. □

Paul Calvert is in the Arizona Materials Laboratories, 4715 East Fort Lowell Road, Tucson, Arizona 85712, USA.

1. Amor, S. R., Rayment, T. & Sanders, J. K. *Macromolecules* **24**, 4583–4588 (1991).
2. Harrison, S. T. L., Amor, S. R., Bonthron, K. M., Chase, H. A. & Sanders, J. K. M. *Int. J. Biol. Macromol.* **14**, 50–56 (1992).
3. de Koning, G. J. M. & Lemstra, P. J. *Polym. Comm.* **33**, 3292–3294 (1992).
4. Lauzier, C., Marchessault, R. H., Smith, P. & Chanzy, H. *Polymer* **33**, 823–827 (1992).
5. Ceccorulli, G., Pizzoli, M. & Scandola, M. *Macromolecules* **25**, 3304–3305 (1992).
6. Barham, P. J. *J. mater. Sci.* **19**, 3829–3834 (1984).
7. Black, S. N., Dobbs, B., Dempsey, P. S. & Daavey, R. J. *J. mater. Sci. Lett.* **9**, 51–52 (1990).

RÉSUMÉ

Frosty reception

WHEN the Roman historian Silius Italicus, describing Hannibal's epic journey over the Alps, wrote of the troops' terror at the sight of the mountains "covered with ice and hail that never thaws", he was taking a rather elastic view of events, it seems. His poetic licence is exposed in *Climatic Change* (22, 139–150; 1992) by J. Neumann, who has taken a close look at tree-ring data for sub-Alpine conifers to show that the climate in 218 BC, the year of the crossing, was if anything milder than in this century. There may have been snow falls in September at the time of the crossing (other historians mentioned this), but no glaciers in the vicinity. But Italicus was not being entirely inventive: during his lifetime, 250–320 years later, the climate had chilled considerably, and the glaciers may have advanced.

The mycoplasma factor

ARE mycoplasmas the elusive cofactors that, with HIV-1, some think are necessary to cause AIDS? Last year S.-C. Lo *et al.* isolated a mycoplasma, dubbed *M. penetrans*, from the urine of AIDS patients. They now show (*Lancet* 340, 1312; 1992) that antibodies to the organism occur in 40% of AIDS patients and 20% of asymptomatic HIV-1-positive patients, but in only 0.3% of uninfected controls (and in none of a group of HIV-1-negative, immunocompromised individuals). As AIDS patients mount poor immune responses, and often do not make antibodies to other mycoplasmas such as *M. fermentans*, the true incidence of *M. penetrans* in AIDS sufferers may be even higher than these figures suggest. Further studies of this organism, using, for instance, PCR of 16S RNA sequences, will no doubt follow.

Weeping wine

OENOPHILES will be no strangers to the sight of droplets running down the inside of a glass of fine wine. Kelvin understood the basis of the effect — differential evaporation of water and alcohol in the film that rises under capillary forces from the edge of the liquid, causing a gradient in concentration and thus surface tension — but the details have been accorded little attention. Now J. B. Fournier and A. M. Cazabat (*Europhys. Lett.* 20, 517–522; 1992) have identified the "tears of wine" as a fingering instability analogous to others, such as the thermal Marangoni effect, which are also the result of surface-tension inhomogeneities. Merging of the tears produces a self-similar profile in the advancing film. As the effect should also be found in engineering operations that involve liquid mixtures with markedly different surface tensions, the results have implications that go beyond the wine cellar.

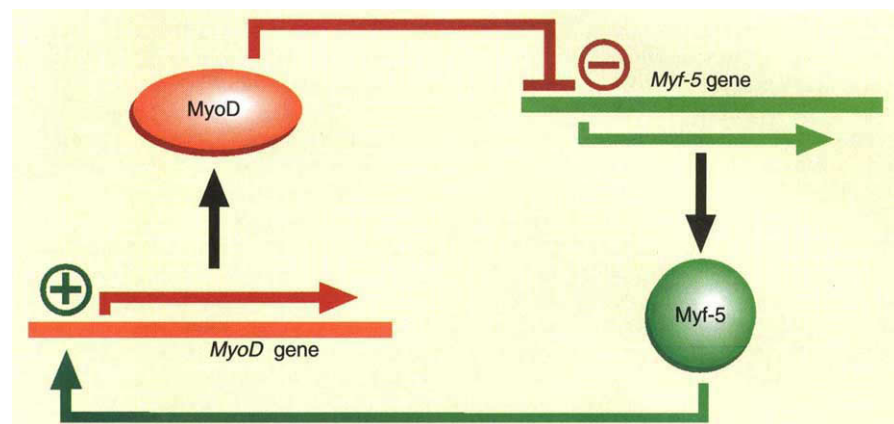
MUSCLE DEVELOPMENT

Running without regulators

Simon M. Hughes

FIVE years ago, the cloning of *MyoD*, a gene that is sufficient to convert fibroblasts into muscle cells, caused quite a stir¹. Many hoped that *MyoD* would prove to be the elusive master-regulator, a switch able to commit a cell to a particular fate and keep it there. The story now turns out not to be so simple, though it may be more intriguing. Two papers in *Cell*^{2,3} show that muscle can form in the absence of either *MyoD*, the original myogenic 'determination' gene,

That the mHLH genes were the key to myogenic commitment then came into doubt when it was discovered that none of their products are detectable in mouse limb buds at a time when muscle precursors are known to be present: could the somitic cells (which populate the limb bud) express mHLH proteins transiently before migration into limbs; do they migrate to the limb before induction of myogenesis; or are other molecules responsible for myogenic commitment?



Possible negative feedback control loop for *MyoD* and *Myf-5*. There is evidence for each separate step, though in different cells. The model is not dependent on whether the mHLH proteins act directly or indirectly on the mHLH genes. Any increase or decrease in either transcription factor activity or gene expression for either *MyoD* or *Myf-5* is immediately self-correcting. Such systems are intrinsically more stable to perturbation than the simple auto-inductive positive feedback loops that have been ascribed to mHLH genes. Together with the myogenin and MRF4 proteins, it is possible that each of several pseudo-stable set points for expression of mHLH genes could maintain a distinct muscle cell state.

or *Myf-5*, the currently favoured candidate in the same gene family.

The *MyoD* gene seemed to be a master-regulator because, in addition to its myogenic activity in cells in tissue culture, it is expressed exclusively in skeletal muscle and activates transcription from the E box DNA motif found in the regulatory regions of most known muscle-specific genes. Moreover, consistent with a switching role, it can activate its own expression by a positive feedback loop. But the unique position of *MyoD* as a master-regulator of myogenesis came under threat when three homologous members of the myogenic helix-loop-helix (mHLH) transcription factor family were discovered. The myogenin, *Myf-5* and MRF4 (also known as *Myf-6*/herculin) proteins each have much the same characteristics as *MyoD* (see ref. 4 for review). The *Myf-5* gene, being expressed before the others, became the favoured putative master-regulator. The question became, do the genes cooperate in a regulatory network or does each have a subtly distinct function *in vivo* (or both)?

Furthermore, experiments in frogs and nematodes showed that some mHLH genes are expressed in cells that are not committed to myogenesis. These clues suggested that the mHLH proteins were unlikely to act alone as master-regulators of myogenesis, so their role had to be tested by manipulations of their expression *in vivo*.

Enter Braun, Rudnicki and their friends^{2,3}, who have modified the *Myf-5* and *MyoD* genes by homologous recombination in embryonic stem cells and produced mice in which, although aberrant transcripts are made, no functional protein should be present. Disappointingly (at least for those who like simple stories), both mice have muscle, so neither gene is the master-regulator of myogenesis. Mice with the *Myf-5^{ml}* mutation die at birth, but not of defective muscles. They suffer from a delayed appearance of somitic muscle accompanied by a failure of rib formation and, hence, an inability to breathe. Braun *et al.*² speculate that the rib defects may be due to failure of an inductive interaction between myogenic tissue and adjacent

sclerotome derivatives during the critical period when muscle is absent. It seems that either myogenic induction is delayed, or that muscle cells are formed but cannot differentiate. Further analysis of both the molecular and cellular features of the mutant is required. *Myf-5^{ml}* does not prevent later expression of other mHLH genes, although a slight decrease in *MyoD* messenger RNA is present in neonates. Perhaps the failure of myogenesis is overcome when one of the other mHLH proteins appears.

Mice with the *MyoD^{ml}* mutation are viable, fertile and apparently normal. However, the survival of homozygous mutants in the first three weeks after birth was less than half that of wild-type or heterozygous litter mates. The reason is unclear but, from the standpoint of evolution, the observation suggests that *MyoD* is not a redundant gene. A second feature of *MyoD^{ml}* is increased expression of *Myf-5*, which is almost undetectable in the wild-type adult. Rudnicki *et al.*³ say that *Myf-5* may functionally substitute for *MyoD*. The phenotype of *MyoD^{ml}Myf-5^{ml}* double mutants will be keenly awaited.

The increase in *Myf-5* expression in *MyoD^{ml}* mice is even more interesting in the light of a decline in *MyoD* mRNA in *Myf-5^{ml}* mutant mice, which is consistent with a host of data on the effects of mHLH genes on one another's expression *in vitro*. It implies that *MyoD* and *Myf-5* may form a negative feedback loop, ensuring that at least one is always expressed (see figure). Yet in early development *Myf-5* is expressed alone, whereas in the adult the *MyoD* protein is much more abundant, suggesting that other factors must influence mHLH expression. Could *Myf-5* and *MyoD* proteins perform the same function in different cell populations? Maybe so — there is no lack of diversity within myoblasts and muscle fibres to account for⁵.

A feature of targeted gene knockouts in mice is that phenotypes are significant, but are often less severe than expected on the basis of the pattern of expression *in vivo* and the properties of the protein in tissue culture cells. This is often put down to redundancy between members of families of homologous genes. But although there may indeed be redundancy between *Myf-5* and *MyoD*, such is not always the case with members of homologous gene families. The dramatic loss of erythrocytes in mice lacking GATA-1 (ref. 6), one member

of a tissue-specific transcription factor family similar to the mHLH family, shows that individual family members can be indispensable for differentiation. There are few correlations between the expression of mHLH proteins and their putative target muscle-specific genes that point to distinct individual functions, one exception being myogenin which is always expressed in myoblasts during terminal differentiation into multinucleate muscle fibres. The knockouts of the genes encoding myogenin and MRF4, currently underway, may not be as anti-climactic as

those of *MyoD* and *Myf-5*.

The fundamental issues of how muscle is formed and how the roles of the mHLH family members differ are still unresolved. It is not clear that the original idea of mHLH proteins as nodal master-regulators is wrong. The question now is, not what are the regulators, but where are the nodes? □

Simon M. Hughes is in the MRC Muscle and Cell Motility Unit and the Developmental Biology Research Centre, King's College London, 26–29 Drury Lane, London WC2B 5RL, UK.

CLUSTERS

Just a fraction of a moment

Tony Stace

BULK ferromagnetism is a cooperative phenomenon associated with unpaired electrons on metal atoms in a crystalline lattice. Each electron behaves as a tiny bar magnet aligned along the spin axis and, although the individual magnetic moments are small, an extended array of

metals are ferromagnetic.

For the same reason, clusters might be expected to have different magnetic properties from bulk materials. Also, the high surface-to-volume ratio of a cluster means that a considerable fraction of the atoms will have coordination numbers

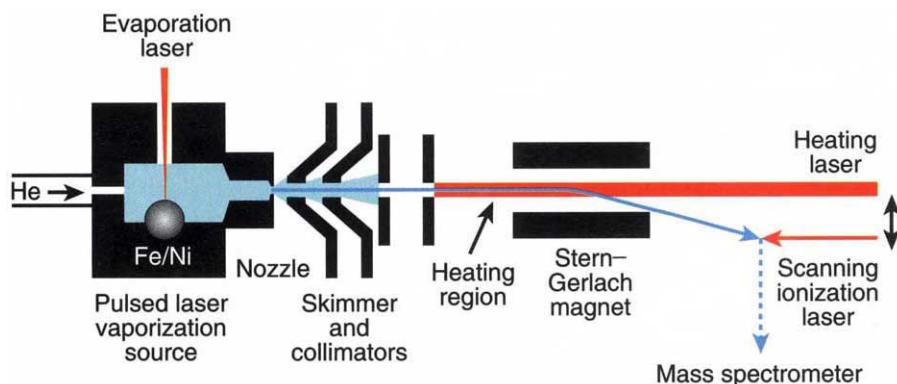


FIG. 1 Apparatus for Becker and de Heer's modified Stern-Gerlach experiment. The entire arrangement is contained within a series of vacuum chambers.

completely ordered moments can generate a large spontaneous magnetization. But what of microscopic clusters, whose properties fall half way between those of molecules and those of bulk materials? Adapting a classic experiment of the 1920s by Stern and Gerlach, Becker and de Heer¹ have measured the magnetism of clusters containing just 100–500 nickel and iron atoms, finding that the standard equations of magnetism do not apply.

Magnetism is frequently described in terms of a model due to Heisenberg², which considers the potential energy between the spins of unpaired electrons on neighbouring atoms. The crucial factor in this model is the strength of coupling between the spins, which is sensitive to the interatomic spacing. This places limitations on the range of interatomic distance suitable for spin coupling and, for this reason, only a few transition

metals are ferromagnetic. Similarly, the absence of an effectively infinite array of atoms means that the modified forces acting on a cluster lattice could result in a change in interatomic distance. Finally, an important consideration in magnetism is temperature; however, this is a quantity that is often difficult to define for an isolated cluster³, in which the important conserved quantity is total energy, not temperature. What influence the restricted heat sink has on spin-lattice relaxation has yet to be explored.

Becker and de Heer's approach has its origins in the famous experiment by Stern and Gerlach in the 1920s. This was decisive in showing that the spin of an electron (and the related magnetic moment) is quantized, so that the magnetic dipole points either 'up' or 'down' with respect to an external magnetic field.

1. Davis, R. L., Weintraub, H. & Lassar, A. B. *Cell* **51**, 987–1000 (1987).
2. Braun, T., Rudnicki, M. A., Arnold, H.-H. & Jaenisch, R. *Cell* **71**, 369–382 (1992).
3. Rudnicki, M. A., Braun, T., Hinuma, S. & Jaenisch, R. *Cell* **71**, 383–390 (1992).
4. Buckingham, M. *Trends Genet.* **8**, 144–149 (1992).
5. Miller, J. B. *Cell* **69**, 1–3 (1992).
6. Pevny, L. *et al.* *Nature* **349**, 257–260 (1991).

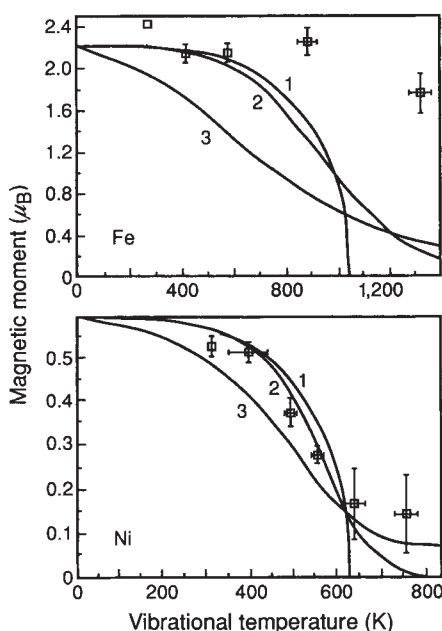


FIG. 2 Magnetic moments per atom (in Bohr magnetons, μ_B) for iron clusters with 120–140 atoms and nickel clusters with 520–600 atoms as a function of cluster temperature. Curves 1 are derived from temperature-dependent data for the two metals in the bulk; curves 2 use the bulk values, but take the heating of the clusters into account; curves 3 are based on the Heisenberg model as applied to clusters.

Stern and Gerlach passed a beam of sodium atoms along the axis between two long magnets, one knife-shaped, the other flat so that the magnetic field between them diverged. The sodium atoms have one valence electron so that their magnetic moment is that of the single electron. The experiment showed that half of the atoms in the beam, those with their dipoles aligned parallel to the field, are drawn towards the knife-shaped magnet (into the higher field), and those with the opposite alignment were pushed away. It was the occurrence of only two paths that revealed the quantization of spin; the degree of the deflection revealed the strength of the dipole.

In Becker and de Heer's experiments, Stern and Gerlach's technique is extended to study clusters of iron and nickel, with the expectation of observing deflections due to the presence of coupled, unpaired spin systems in the form of microscopic ferromagnets. The clusters were made by evaporating metal samples with high-powered laser pulses, and entraining and cooling the vapour in a stream of helium, in which the clusters formed by condensation (Fig. 1). Gas and clusters were then forced through a supersonic expansion, in which there is further cooling, and passed down the magnet axis. The idea was to find the cluster magnetization, $M_z = N\mu^2 H/3kT$ (μ is the intrinsic dipole of each atom, H

the applied field and k Boltzmann's constant) as a function of cluster size (N , number of atoms) and temperature T . The deflection of the clusters was determined by ionizing them with a second laser, which could be scanned across the face of the magnets, and recording their arrivals in a mass spectrometer. In contrast with the original single-atom experiment, the cluster experiment gives an asymmetric distribution of deflections, towards high fields, an effect attributed¹ to the way coupling between the atoms allows them to exchange energy and angular momentum, something not possible with an isolated atom.

The magnetization could be determined from the beam deflection. The cluster temperature was derived from the rate at which atoms evaporated from the cluster following heating by yet another laser, using a formulation developed by Klots five years ago (see my previous News and Views article³). In this way, Becker and de Heer could compare the magnetization of the clusters with that of bulk materials (Fig. 2). At low temperatures, the cluster magnetic moments are close to those found for each of the bulk metals: for iron, $\mu_{\text{bulk}} = 2.2$ Bohr magnetons; for nickel, $\mu_{\text{bulk}} = 0.6$ Bohr magnetons. As T increases towards the Curie temperature, the magnetic moment of a bulk metal rapidly approaches zero; similar behaviour is seen in the nickel clusters, but not for those of iron. It is clear from the results that the behaviour of the latter follows neither the Heisenberg model nor the temperature dependence of a bulk ferromagnet.

Becker and de Heer are not the only group to be active in this area^{4,5} and, although their analysis of Stern and Gerlach's experiments on small metal clusters seems to be the most comprehensive to date, there are still a number of unresolved problems in the interpretation. Experiments on metal clusters offer opportunities for understanding magnetic behaviour at an atomic level; however, the experiments could also help to identify new ferromagnetic materials. An example here might be manganese; with the metal in the form of small clusters, lattice relaxation and/or a change in lattice structure could increase the interatomic coupling to a point where, over a narrow size range, the clusters become ferromagnetic. □

Tony Stace is in the School of Chemistry and Molecular Sciences, University of Sussex, Falmer, Brighton BN1 9QJ, UK.

1. Becker, J. A. & de Heer, W. A. *Ber. Bunsenges. Phys. Chem.* **30**, 1237–1243 (1992).
2. Blakemore, J. S. *Solid State Physics* (Cambridge University Press, 1985).
3. Stace, T. *Nature* **327**, 186–187 (1987).
4. Cox, D. M. et al. *Phys. Rev. B* **32**, 7290–7298 (1985).
5. Bucher, J. P. et al. *Phys. Rev. Lett.* **66**, 3052 (1991).

DAEDALUS

The iron horse

AGRICULTURE is, theoretically, a way of capturing the Sun's energy for human use by photosynthesis. Modern 'efficient' agriculture in fact runs at a huge loss. The energy used to make fertilizers, drive tractors, process and pack the products and get them to market, far exceeds the solar energy thus captured. Even ecologically correct schemes like converting corn into fuel-alcohol, or rape oil into diesel fuel, usually run at an energetic loss. They are simply an expensive massage for the industrial conscience.

So Daedalus has been dreaming up a genuinely energy-efficient agriculture. He is devising a diesel engine that runs directly on plant material and returns its waste to the soil, as a horse does. He recalls that Rudolf Diesel's original engine was designed to burn coal dust. DREADCO engineers are now developing a diesel to burn powdered leaves, corn-stalks, rice-hulls and so on.

The easiest way to pulverize biological materials is to embrittle them by cooling with liquid air and to grind them up. All internal combustion engines sacrifice some of their power to drive a cooling system; the DREADCO cryodiesel uses a significant fraction of its power in this way, to operate its on-board cryogenic plant. The resulting flow of cryogenically cold powder is squirted directly into its cylinders.

Very fine powders burn extremely efficiently. In any case, poor burning can be neatly overcome with the aid of liquid oxygen generated by the cryogenic system. The resulting exhaust will contain a lot of solid ash, but with luck this will be so finely divided that it will polish the working parts of the engine, rather than abrading them. As a final neat touch, the waste cold from the cryogenics will be used to cool the engine's exhaust gas to ambient temperatures. The steam of combustion will condense to water, in which the alkaline plant ash will be trapped and suspended. The acidic nitrogen oxides of combustion will react with the suspension to give a useful nitrogenous fertilizer slurry. Instead of generating an unpleasant smoky exhaust, the new engine will return its waste usefully to the land.

Thus modern wasteful agriculture will be reformed. Even old-style agriculture was inefficient, in the sense that the greater part of any crop plant cannot be eaten by men or animals. But Daedalus's scheme uses all the inedible and waste material. Instead of being rejected by dumping or burning, these biological wastes will be used as fuel, and their mineral contents recovered as fertilizer. Even modern industrial farming waste, such as plastic bags and old tyres, will be reclaimed. David Jones

Oak tree mortality in Iberia

SIR — Rapid death and decline of the cork oak *Quercus suber*, *Q. ilex* and other oak species is occurring across parts of the Mediterranean. In 1991 more than 1,050 major decline foci were recorded in the approximately 2.2 million hectares of oak forests and plantations in southwest Spain (Fig. 1), including 265 foci of average 21.4 ha in the 0.1 million ha of oak cover in the Parque Natural de Alcornocales (natural oak park) in Andalucía. About half the trees in a focus were dead or dying (R. Montoya, ICONA Madrid, personal communication). Comparable oak decline is occurring in adjacent areas of southern Portugal (Fig. 1), and in Italy, Morocco and Tunisia. As in the still unexplained episodes of oak mortality across central and eastern Europe since about 1900 (ref. 1), Mediterranean oak decline has so far been attributed mainly to drought, pollution and to secondary attacks by insects and fungi.

Between May 1991 and March 1992, at the invitation of the Ministry of Agriculture, Madrid and the University of Algarve, Faro, I examined a range of sites in Iberia where oak had declined. Both tree symptoms and decline distribution suggested the possible spread of a root disease caused by the soil- and water-borne fungus *Phytophthora*. Trees often died suddenly in one or two seasons, and tarry exudations and epicormic shoots indicated root stress. Dying trees often occurred in groups, and large foci were noticeably distributed along streams, valleys or depressions (Fig. 2). Rapid decline was sometimes associated with winter surface-lying water or with

recent soil disturbances such as ploughing or road making. A more chronic decline occurred on the dryer hillsides. Maquis shrubs (for example, *Cistus* and *Lavendula*) were also dying on some sites.

On dryer soils, dying trees often had many dead fine feeder roots. On more moist soils some trees also had dead and dying larger roots with necrotic bark lesions. Using apple baits and selective antibiotic media, the highly aggressive exotic root pathogen *Phytophthora cinnamomi* was isolated from roots of affected *Q. suber* and *Q. ilex*, or from associated soil, at eleven of thirteen decline sites investigated by myself and local colleagues (Fig. 1)². *P. cinnamomi* is ephemeral and difficult to isolate under dry conditions³. Most positive isolations came from sites where soil had remained moist, and our failure to recover the fungus from two sites may reflect the very dry soil conditions due to recent unseasonal droughts.

P. cinnamomi is an unusually polyphagous pathogen of mainly woody hosts⁴, most active in roots at 25–30 °C (ref. 3). Probably indigenous to the New Guinea-Celebes region, it has become widely distributed by man. Its introduction is considered responsible for recent dieback of entire eucalypt forest communities in western Australia and for the mass dying of native chestnuts across the south-eastern United States in the early 1900s (refs 3, 4). It was becoming widespread in Europe by the 1940s, causing a major epidemic on the European chestnut⁵. From tree symptoms and from its close association with decline sites, the spread of *P. cinnamomi* is likely to be a major factor in the rapid oak mortality in Spain and Portugal, interacting with winter droughts (a recent climatic feature of affected areas) and with changing land use patterns. This may lead to secondary stress attacks by insects and other fungi. By anal-

ogy, *P. cinnamomi* may also be involved in the similar oak declines elsewhere in the Mediterranean.

In addition to its intrinsic significance, the association with *P. cinnamomi* highlights once again the threat posed to forest ecosystems by the introduction of exotic pests or pathogens: a threat which



FIG. 2 Rapidly dying *Q. ilex* in a depression (foreground) and more healthy *Q. ilex* on a hillside (background) at an oak decline site in western Spain. *P. cinnamomi* was isolated from roots of *Q. ilex* at this site.

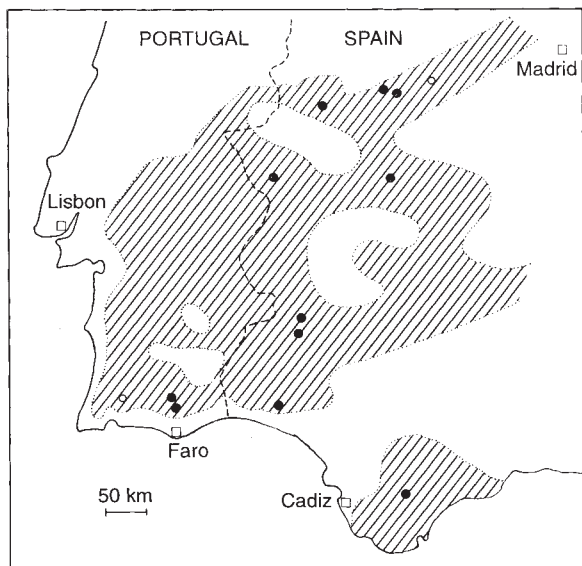


FIG. 1 Main distribution of *Q. suber* and *Q. ilex* (hatched area) in southwest Spain and southern Portugal. ●, Decline site at which *P. cinnamomi* isolated; ○, decline site at which *P. cinnamomi* not isolated. For details see ref. 2.

in recent years has been overshadowed by the issues of atmospheric pollution and forest exploitation. Such introductions frequently involve trade, and it is exceedingly difficult to ensure that adequate attention is given to the environmental risks posed by the increasing diversity of world trade, and to safeguard the resulting intercontinental movement of plants, raw materials and manufactured goods. Another concern is that *P. cinnamomi* may represent the sort of organism that could become more damaging under conditions of general climatic warming, either directly through becoming more active at higher temperatures, or indirectly through exploiting stress effects on host vegetation. At present these risks to forest ecosystems cannot reliably be predicted or assessed.

C. M. Brasier
Forest Research Station,
Alice Holt Lodge,
Farnham,
Surrey GU10 4LH, UK

1. Oleksyn, J. & Przybyl, K. *Eur. J. Forest Path.* **17**, 321–326 (1987).
2. Brasier, C. M., Robredo, F. & Ferraz, J. F. P. *Plant Path.* (in the press).
3. Shearer, B. L. & Tippet, J. T. *Department of Conservation and Land Management, Como, Western Australia Research Bulletin* **3**, 1–76 (1989).
4. Zentmyer, G. A. *Am. Phytopath. Soc. Monogr.* **10**, 1–96 (1980).
5. Moreau, C. & Moreau, M. *Rev. Path. veg. Ent. agric. Fr.* **31**, 201–231 (1952).

Palaeocene therapsid debate

SIR — Sues¹ contests our identification of *Chronoperates* as a Palaeocene mammal-like reptile of the Order Therapsida², an identification which extends the temporal range of therapsids over 100 million years. Sues implies that *Chronoperates* is instead a mammal, but presents no synapomorphies that it shares with mammals, nor does he tell us what he conceives a mammal to be. The dimensions specified for the Class Mammalia differ widely and, because we do not know which alternative Sues prefers, we are unable to evaluate the features of *Chronoperates* that he thinks are mammalian. Sues, following Novacek³, suggests that the differences between the teeth of *Chronoperates* and symmetrodonts (early therian mammals restricted to the Mesozoic) are 'subtle'. However, neither of them has accounted for the lack in *Chronoperates* of several fundamental features that are seen in the cheekteeth of all symmetrodonts (for example, double roots, anterior peg and posterior recess interlocking mechanism, basal cingulum or well-defined shearing surfaces).

Sues identifies several features of *Chronoperates* which demonstrate that it is not one of several already known Triassic cynodonts, and we agree completely. Nonetheless, we still maintain, and Sues has not shown otherwise, that the structure of the teeth and dentary of *Chronoperates* is more similar to that of Triassic cynodonts than to that of any other known vertebrate, including any Mesozoic mammal. Further, Sues states that we distinguished between *Chronoperates* and Mesozoic mammals on differences in enamel microstructure and were incorrect in doing so; however, our comparison was not with Mesozoic mammals, but with two Cenozoic eutherian groups, Edentata and Mesonychia, both of which have teeth which superficially resemble those of *Chronoperates*. As we stated originally, the enamel of eutherians is fully prismatic, not pseudo-prismatic as in *Chronoperates*.

Sues implies, misleadingly, that the therapsid features of the dentary of *Chronoperates* are probably artefacts of preservation and not original anatomical structures. In reply, we simply note that Sues has never observed the original specimens or casts of them, with or without magnification. Although breakage and crushing may obscure demonstration of a shallow Meckelian groove

(as we explicitly noted), there is no evidence that the other features in question (such as the presence of a groove for postdentary bones and scar for coronoid bone, inferior limit of masseteric fossa) are artefacts resulting from breakage, crushing or some combination that affected the specimen after death. We therefore maintain that our original interpretations were correct and invite Sues and others to examine the specimens. Our only deficiency seems to be our failure to interpret the evidence according to Sues's expectation of what it 'should' be. In concluding from the

evidence that *Chronoperates* is a Palaeocene therapsid, we think we have done substantially better than that.

Richard C. Fox

Gordon P. Youzwyshyn

Laboratory for Vertebrate Paleontology,
Departments of Geology and Zoology,
University of Alberta,
Edmonton, Alberta T6E 2E9,
Canada

David W. Krause

Department of Anatomical Sciences,
State University of New York,
Stony Brook,
New York 11794-8081, USA

Jury returns on structure prediction

SIR — The jury on structure prediction has returned. The witnesses: the experimental structure of the Src-homology 3 (SH3) domain by Musacchio *et al.*¹, solved by X-ray crystallography, and the accompanying theoretical structure by Benner *et al.*², predicted from the amino-acid sequence. The case: how well did the ETH (see figure) structure-prediction method work in this blind test, performed before the structure became available?

First, the bad news: the tertiary structure prediction was wrong. Benner *et al.*²

structure, the answer is yes: our novel prediction method³, also tested on the SH3 domain without knowledge of the structure, reached not only 80% for segments, but also 70% for single residues, a good result by current standards. This new method (PHD in the figure) uses information from multiple sequence alignments (as does the ETH method), but has the added advantage of being fully automatic (the ETH method relies in part on human intuition).

We agree with Benner *et al.* that blind predictions should be made before ex-

	1.....2.....3.....4.....5.....								
AA	KELVLALYDYQEKSPREVTMKKGDIITLLNSTNKDWWKVEVNDRQGFVPAAYVKKLD								
EXP	EEEE B B B EEEEE EEEEE EEEEEEGGEEEE								
ETH	EEEEEE EEEEE HHHHHHHH EEEE EEE EEEEE								
PHD	EEEEEE EEE EEEEE HHHHHH EEEE HEEE								
MUS	EEEEEE EEEEE EEEEE EEEEE EEEEEEE								

Comparison of the secondary structure of the SH3 domain, determined experimentally and predicted on the basis of multiple sequence alignments by three different methods. AA, amino acid sequence; EXP, crystal structure¹; ETH, *de novo* prediction of conformation²; PHD, profile neural network prediction (jury of 12 networks)³; MUS, partial prediction by Musacchio *et al.*⁵; E, β -strand (extended); H, α -helix; G, 3_{10} -helix; B, isolated β -bridge.

predicted a structure "built from β -strands with a single turn of α -helix lying on one face". The crystallographers¹, instead, find that "the five strands form two orthogonal β -sheets, as in a β -sheet sandwich". The only similarity between the predicted and experimental tertiary structures is the presence of β -strands, and that is merely a statement about secondary structure.

Second, the good news: the secondary structure prediction was quite accurate. The protein is correctly predicted to consist mainly of β -strands. Of the five β -strands, four are correctly predicted in about the right sequence position, and one is predicted incorrectly as a helix, a neat 80% success rate for segments (ETH in the figure). The success rate is less good when counting how many residues are predicted correctly as helix, strand or loop: only 56%.

Can one do better? For secondary

experimental structure solution. To make such tests independent of human intervention, we have installed an electronic mail server to which academic researchers can submit sequences (send 'help' to PredictProtein@Embl-Heidelberg.De on Internet). The secondary structure prediction will be sent by return mail. With the steady stream of structures now appearing almost weekly, we will know in a year's time the true accuracy of "prediction of progress at last"⁴.

Burkhard Rost

Chris Sander

Protein Design Group,
European Molecular Biology Laboratory,
D-6900 Heidelberg, Germany

1. Sues, H.-D. *Nature* **359**, 278 (1992).

2. Fox, R. C., Youzwyshyn, G. P. & Krause, D. W. *Nature* **358**, 233–235 (1992).

3. Novacek, M. J. *Nature* **358**, 192 (1992).

1. Musacchio, A. *et al.* *Nature* **359**, 851–855 (1992).

2. Benner, S. A., Cohen, M. A. & Gerloff, D. *Nature* **359**, 781 (1992).

3. Rost, B. & Sander, C. *J. Neural Syst.* (in the press).

4. Thornton, J. M., Flores, T. P., Jones, D. T. & Swindells, M. B. *Nature* **354**, 105–106 (1992).

5. Musacchio, A. *et al.* *FEBS Lett.* **307**, 55–61 (1992).

Ferroelectricity in oxides

SIR — Cohen¹, using total energy calculations, ascribes the origin of ferroelectricity in BaTiO₃ and PbTiO₃ to transition metal–oxygen *p*–*d* hybridization effects. Here, we emphasize that this approach is, of course, not unique because dynamical properties, which are very important in ferroelectrics, are neglected by Cohen.

A unified interpretation of ferroelectricity that is not restricted to perovskite oxides has been offered by Migoni *et al.*², who argued that the O²⁻ ion may be crucial for the polar ferroelectric state to occur. These authors took into account the outstanding properties of O²⁻ by introducing a shell model description for the lattice dynamics of perovskite oxides. The temperature dependence of the oxygen-ion polarizability, which is assumed to drive the lattice instability, has been taken into account by introducing a fourth-order repulsive term in addition to a local attractive electron–ion coupling, thus inducing a local double-well potential in the core–shell interaction at the oxygen-ion lattice site. Within self-consistent phonon theory, quantitative agreement with experimental data has been achieved for phonon dispersion relations, second-order Raman spectra, temperature dependence of the soft-mode frequency and coupled branches as well as defect

properties.

Extensions to systems other than perovskite oxides have been carried out within a simplified version³ of the model of Migoni *et al.*, which quantitatively describes the temperature dependence of soft modes in various structurally different compounds⁴. The extremely transparent, physically understandable and analytically tractable model possesses, besides a wide range of applicability, highly interesting nonlinear solutions which exist on the lattice as well as in a continuum approach to the lattice⁵. From these studies on ferroelectric systems, we conclude that the origin of ferroelectricity in oxides is a consequence of the critical dynamics which are driven by crucial interplay of ionic and electronic interactions.

A. Bussmann-Holder
Max Planck Institut für
Festkörperforschung,
Heisenbergstr. 1,
D-7000 Stuttgart 80,
Germany
H. Büttner
Institut für Physik,
Universität Bayreuth,
D-8580 Bayreuth,
Germany

1. Cohen, R. E. *Nature* **358**, 136 (1992).
2. Migoni, R., Bilz, H. & Bäuerle, D. *Phys. Rev. Lett.* **37**, 1155 (1976).
3. Bilz, H., Benedek, G. & Bussmann-Holder, A. *Phys. Rev. B* **35**, 4840 (1982).
4. Bussmann-Holder, A., Bilz, H. & Benedek, G. *Phys. Rev. B* **39**, 9214 (1989).
5. Bussmann-Holder, A., Büttner, H. & Bilz, H. *Solid State Physics* **48** (ed. Turnbull, D.) (Academic, Ehrenreich, in the press).

Deviant TATA-box binding protein

SIR — In their paper reporting the crystal structure of one variant of the *Arabidopsis* TATA-box binding protein (TBP), Nikolov *et al.*¹ draw attention to the remarkable similarity of the C-terminal domains of all known TBP proteins, as does Greenblatt in the accompanying News and Views article². Thus, in this 180-residue region, TBP molecules of fungal, plant, insect and mammalian origin are a minimum of 70% identical at the amino-acid level. On this basis, and the fact that variant residues map to positions unlikely to perturb the folding pattern, it is suggested that the reported structure of the TBP DNA-binding domain may be universal.

We have characterized the TBP homologue from the human malaria parasite *Plasmodium falciparum*³, which we find to be strikingly divergent from these examples. In this case, the corresponding amino-acid identities range from 38 to 44%. In this organism, whose genome shows a highly biased base composition, putative promoter regions have an A+T

content averaging close to 90% (ref. 4). Under these circumstances, the unambiguous recognition of conventional TATA-boxes will be difficult, and the observed divergence in protein composition may reflect an evolutionary response to this problem.

In any case, this dramatically altered sequence composition means that caution is needed in regarding the *Arabidopsis* TBP structure as being universal. In particular, the *P. falciparum* molecule may be subtly different from the structure presented. Any such differences between it and its equivalent from the human host might well be reflected in functional differences that in future could be exploited therapeutically.

We would also point out that the residual sequence identities conserved between the *P. falciparum* TBP protein and those of other organisms (57, compared to the 126 seen in the data set of Nikolov *et al.*¹) may help to define a much reduced number of residues likely to be critical to the interaction of TBP

with other members of the transcription initiation complex. This information could aid the informed design of site-directed mutagenesis experiments to investigate further the nature of intermolecular interactions in other eukaryotic transcription systems.

John E. Hyde
Paul F. G. Sims
Martin Read
Michael B. McAndrew
Department of Biochemistry and Applied
Molecular Biology,
UMIST, PO Box 88,
Manchester M60 1QD, UK

1. Nikolov, D. B. *et al.* *Nature* **360**, 40–46 (1992).
2. Greenblatt, J. *Nature* **360**, 16–17 (1992).
3. McAndrew, M. B. *et al.* *Gene* (in the press).
4. Hyde, J. E. and Sims, P. F. G. *Gene* **61**, 177–187 (1987).

More on motor neurons

SIR — Vrbová *et al.*¹ raise questions with regard to our paper on the effect of ciliary neurotrophic factor (CNTF) on the mouse mutant progressive motor neuronopathy (*pmn*)². They wonder whether the “relatively modest loss of facial motor neurons” (40%) can account for the death of untreated homozygotes. It certainly cannot.

However, the denervation of the diaphragm at the fourth postnatal week caused by the almost complete loss of motor axons in the phrenic nerve³ would be sufficient to lead to the death of the animals 2 weeks later. The putative “sharp contrast” they note between the original observations³ and our manuscript², that motor neurons of the spinal cord and cranial nuclei appear qualitatively normal at 5 weeks of age³ whereas 40% of the facial motor neurons have degenerated at 6–7 weeks², simply reflects the rapid progression of the degeneration process in the *pmn* mutant. Moreover, cell numbers of spinal motor neurons were not determined in the original paper³ so that loss of these neurons cannot be excluded. Similar observations (no loss of anterior horn cell somata) have also been reported for other animal models such as the Brittany spaniels with hereditary canine spinal muscular atrophy (HCSMA)⁴. If one should follow the criterion of Vrbová *et al.*, that any useful model for motor neuron disease should show an initial degeneration of the motor neuron cell body, then the HCSMA model should also be dismissed.

The *pmn* mutant is certainly not identical with amyotrophic lateral sclerosis (ALS) (we never claimed that), but the pathological manifestations of ALS show striking similarities to those observed in

pmn mice. The traditional dogma that loss of motor neuron somata represents the primary lesion in ALS has never been supported by convincing data; on the contrary, it has been seriously questioned for both ALS⁵⁻⁸ and spinal muscular atrophy⁹. Fewer myelinated nerve fibres are present in distal compared to proximal segments of the phrenic nerves of ALS patients¹⁰, arguing in favour of a 'dying back' process as observed in *pmn* mice. Moreover, the pronounced axonal degeneration and regeneration in the ventral roots of ALS patients¹¹ is not compatible with the assumption of primary degeneration of motor neuron somata. Indeed, electron micrographs of degenerating motor endplates⁷ strikingly resemble the initial changes detected in *pmn* mice.

We conclude from our data that CNTF not only prevents the loss of motor neuron somata, but also maintains axonal integrity and supports axonal regeneration. A recent report on the effects of CNTF on sprouting of motor neurons from endplates and nodes of Ranvier¹² is in agreement with this interpretation. The consequence of these effects is functional restoration of the motor innervation of the paralytic skeletal musculature. Thus, that CNTF interferes with pathological manifestations at both functional and morphological levels appears relevant to the potential usefulness of CNTF in the treatment of ALS patients.

Michael Sendtner

Kurt A. Stöckli

Patrick Carroll

Georg W. Kreutzberg

Hans Thoenen

Max-Planck-Institute for Psychiatry,

D-8033 Martinsried,

Germany

Henning Schmalbruch

Department of Neurophysiology,

Panum Institute,

University of Copenhagen,

DK-2200 Copenhagen,

Denmark

1. Vrbová, G., Greensmith, L. & Sieradzan, K. *Nature* **360**, 216 (1992).
2. Sendtner, M. et al. *Nature* **358**, 502-504 (1992).
3. Schmalbruch, H. et al. *J. Neuropath. exp. Neurol.* **50**, 192-204 (1991).
4. Cork, L. C. et al. *Lab Invest.* **61**, 69-76 (1989).
5. Hirano, A. et al. in *Motor Neuron Diseases* (eds Norris, F. H. & Kurland, L. T.) 51-60 (Grune & Stratton, New York, 1969).
6. Ghatak, N. R. et al. *J. Neuropath. exp. Neurol.* **45**, 385-386 (1986).
7. Festoff, B. W. in *Amyotrophic Lateral Sclerosis* (ed. Davidoff, R. A.) Handbook of the Spinal Cord Vol. 4, 607-684 (Dekker, New York, 1987).
8. Festoff, B. W. in *Handbook of Amyotrophic Lateral Sclerosis* (ed. Smith, R. A.) 661-685 (Dekker, New York, 1992).
9. Hausmanowa-Petrusewicz, I. in *Spinal Muscular Atrophy Infantile and Juvenile Type*, 1-180 (National Center for Scientific, Technical and Economic Information, Warsaw, 1978).
10. Bradley, W. G. et al. *Ann. Neurol.* **14**, 267-277 (1983).
11. Hanyu, N. et al. *J. neurol. Sci.* **55**, 99-115 (1982).
12. Gurney, M. E., Yamamoto, H. & Kwon, Y. *J. Neurosci.* **12**(8), 3241-3247 (1992).

Per — no link to gap junctions

SIR — It has recently been determined that the results of certain experiments originally performed in the laboratory of one of the authors (D.C.S.), and described in our paper¹ entitled "The *Drosophila* clock gene *per* affects intercellular junctional communication", cannot be reproduced. This paper reported changes in gap-junction-mediated intercellular communication in salivary glands from *D. melanogaster period* (*per*) mutants. Three classes of coupling measurements were performed in the original study: (1) measurement of dye spread among cells in *per*⁰, *per*^s and *per*⁺ salivary glands; (2) determination of electrotonic coupling in intact salivary glands of these genotypes; and (3) determination of electrical coupling between isolated cell pairs from dissected salivary glands of these genotypes. For all three classes of experiments, measurements were performed by D.C.S. on material supplied by T.A.B. and M.K.B. This work led to a central conclusion put forth in the paper — that by all three criteria strong differences in intercellular coupling distinguished the *per* genotypes.

A reinvestigation of dye coupling² has concluded that no consistent differences in intercellular coupling in salivary glands can be attributed to *per* locus mutations. This new report² was initiated by K. Flint, M. Rosbash and J. C. Hall, who communicated their unpublished results to us. Similar data were then obtained by D.C.S. and K. Siwicki, and these two sets of data were pooled. The new paper, and the absence of any indication as to the cause of the discrepancy between the two studies, causes us to question the validity of the original measurements of dye and electrical coupling. We also wish to point out that these data were essential for support of a model proposed in our paper: that *per* might control circadian rhythms by modulating intercellular junctional communication in the nervous system.

In addition to the studies of intercellular communication described above, our paper included data supporting two other conclusions: (1) that antibodies against the *per* product can detect an antigen of unexpectedly high molecular mass, which is reduced in size by treatment with heparinase II; and (2) that *per* expression can be detected in developing *Drosophila* salivary glands. These experiments were carried out in the laboratory of M.W.Y. We note that our detection of antigen with proteoglycan properties was consistent with an earlier, related biochemical study of *per* antigens

by others³, and we know of no studies that further directly address this issue. We have no reason to question the data presented in support of this aspect of our study. Indirect evidence does suggest that such modification of *per* may not be required for expression of circadian rhythms⁴. As has been previously acknowledged⁵, *per* expression in developing salivary glands has not been confirmed by studies from another laboratory^{6,7}, while, for example, work in M.W.Y.'s laboratory indicates consistent labelling of these glands with *per* antibodies and with strand-specific *per* RNA probes following *in situ* hybridization. More direct support for *per* expression in developing salivary glands has come from blot analysis of RNA from hand-dissected larval tissue⁸. The latter analysis also shows that hybridizing transcripts from the glands are indistinguishable in size from *per* RNA extracted from heads⁸. It may also be important that evidence for *per* expression in *adult* salivary glands has been presented by others⁶.

In conclusion, we wish to retract those portions of our paper reporting evidence for the involvement of *per* in gap-junction-mediated intercellular communication. We acknowledge the shared responsibilities for that work as coauthors of a collaborative effort. We deeply regret any difficulties our original report may have caused other researchers working in this field.

Lino Saez

Michael W. Young*

The Rockefeller University

*Howard Hughes Medical Institute,

New York,

New York 10021, USA

Mary K. Baylies

Department of Zoology,

University of Cambridge,

Cambridge, CB2 3EJ, UK

Gregory Gasic

Salk Institute,

La Jolla,

California 92037, USA

Thaddeus A. Bargiello

David C. Spray

Department of Neuroscience,

Albert Einstein College of Medicine,

Bronx,

New York 10461, USA

1. Bargiello, T. A. et al. *Nature* **328**, 686-691 (1987); correction *Nature* **360**, 610 (1992).
2. Siwicki, K. K., Flint, K., Hall, J. C., Rosbash, M. & Spray, D. C. *Biol. Bull.* (in the press).
3. Reddy, P., Jacquier, A. C., Abovich, N., Petersen, G. & Rosbash, M. *Cell* **46**, 53-61 (1986).
4. Rosbash, M. & Hall, J. C. *Neuron* **3**, 387-398 (1989).
5. Saez, L. & Young, M. W. *Molec. cell. Biol.* **8**, 5378-5385 (1988).
6. Liu, X., Lorenz, L., Yu, Q., Hall, J. C. & Rosbash, M. *Genes Dev.* **2**, 228-238 (1988).
7. Siwicki, K. K., Eastman, C., Petersen, G., Rosbash, M. & Hall, J. C. *Neuron* **1**, 141-150 (1988).
8. Baylies, M. K., Weiner, L., Vossball, L. B., Saez, L. & Young, M. W. in *Molecular Genetics of Biological Rhythms* (ed. Young, M. W.) 123-154 (Dekker, New York, in the press).

Glimpses of the past

William A. S. Sarjeant

Scenes From Deep Time: Early Pictorial Representations of the Prehistoric World. By Martin J. S. Rudwick. University of Chicago Press: 1992. Pp. 280. \$45, £39.95.

WE live in a period of unprecedented richness, in terms of our knowledge of the past and our ability to portray it vividly. In unearthing and interpreting the evidences of past times, geologists, palaeontologists, archaeologists and prehistorians have helped us to understand how Earth evolved through countless aeons and how our culture developed from the first shaping of artefacts to the infinite complexities of our modern civilization. Through the inventiveness of artists and engineers, we can learn of that past not only from paintings and sculptures but also from three-dimensional dioramas, moving pictures on small and large screens, working models and the new wonder of holograms.

In particular, the events and the denizens of the geological past are now so freely available to us that they have become almost commonplace. As Rudwick states of dinosaurs: "... at any natural history museum today ... [w]e see their reconstructed skeletons, often rearing above our head; we see more lifelike reconstructions of their whole bodies, skin and all; and now we may even see them heaving from side to side, clamping their jaws and grunting, thanks to the magic of modern robotics."

This book shows how the first steps were taken along the road that led from the first flowering, at the beginning of the nineteenth century, of the study of fossils, to our present rich understanding and vivid representations. Rudwick argues reasonably that the earliest palaeontological reconstructions were derived from the pictorial representations of the much later — though to the artists, still distant — periods of the Bible. Without doubt, the legends of the origins of Earth set forth in Genesis conditioned the thinking of early geologists; such biblical scenes would have been familiar to them from their earliest

days in church and school. Moreover, as Rudwick reveals, it was the German palaeontologist J. J. Scheuchzer who, in his *Physica Sacra* (1731–33), first sequentially illustrated episodes in scriptural history from Genesis to the apoca-



Reptilian dreams — "Plesiosaur", the first scene from deep time in Pierre Boltard's *Paris Before Men* (1861). The narrator has been transported by the magic of his companion, the 'lame demon'.

lypse, anticipating the method by which geological periods are nowadays represented.

The first reconstruction of a fossil vertebrate skeleton was made in about 1790 in Madrid, where J. B. Bru successfully reassembled the bones of the extinct giant ground-sloth *Megatherium*. It was the great French anatomist Cuvier, however, who first determined the musculature of an extinct creature, the Eocene mammal *Anoplotherium*, and sketched its bodily outline, while two British palaeontologists, W. D. Conybeare and H. T. de la Beche, were the first to fill in such outlines and place the

reconstructed creatures against scenic backgrounds. Both had nonscientific aims; Conybeare's caricature of William Buckland entering a den of cave hyenas accompanied some humorous verses about his friend, while de la Beche's spirited representation of the teeming life of Jurassic seas was a benefaction to the Dorset fossil collector Mary Anning.

Indeed, as Rudwick demonstrates, the early palaeontologists were reluctant to exercise their imaginations in this way, fearing it might prove prejudicial to their reputation for scientific probity. In consequence, developments from those beginnings are found not in serious studies by renowned savants, but in popular or eccentric works by people on the scientific fringe. When the Austrian botanist F. X. Unger commissioned an artist to produce a sequence of scenes representing the changing faunas and floras of successive geological periods, he did so reluctantly, afraid of "wandering from the actual domain of science, into the realms of imagination". Only with Waterhouse Hawkins's reconstructions of dinosaurs, made under Richard Owen's supervision for the Crystal Palace exhibition, was this particular bogey laid to rest.

Rudwick has produced a thoroughly researched history of the development of pictorial palaeontological reconstruction up to the influential work of Louis Figuier, and his artist Edouard Riou, in the 1860s. The numerous figures portray with admirable clarity the evolution of this discipline, a fascinating interplay of science with art. With the illustrations and research, I have no quarrels, although the

text itself occasions a number of quibbles. First, the names of scientists are cited inconsistently. The full name of the Swiss scientist Pictet is presented (p. 88), whereas that of Jacques Boucher de Crèvecœur de Perthes is curtailed (p. 166) and the resounding Christian names of Cuvier, only called 'Georges' after the death of his elder brother, are not given at all (p. 30). Second, I can well comprehend Rudwick's decision not to indicate the generic names now in use for fossil genera mentioned, but surely readers would have been interested by such matters as the early misplacement of *Iguanodon*'s spiked thumb as a nasal

From Scenes From Deep Time

horn, the misunderstanding of the manner of flight of what Rudwick terms "pterodactyles", and the lack of awareness of the dorsal fin and tail-shape of ichthyosaurs. I must fault him for his confusion of *Cheirosauros*, *Chirotherium* and *Labyrinthodon* — the first two genera based on footprints, the last on bones — in his commentaries on Figures 47 (where the beast in the foreground gains no mention in the text), 49, 86 and 87. (These footprints were produced not by a quadrupedal amphibian, but by a bipedal reptile). Indeed, several vertebrates fare badly. *Archegosauros* is not "the earliest reptile" (p. 180) but an

amphibian; *Rhamphorhynchus* is misspelled (p. 190); and the aurochs was not an "extinct wild bison" (p. 128) but the direct ancestor of our modern cattle.

Despite these carpings, this is a book that I can recommend wholeheartedly — an excellent concept, excellently executed. Historians of science and art, and indeed anyone interested in the creatures of the past, should find it fascinating. □

William A. S. Sarjeant is in the Department of Geological Sciences, University of Saskatchewan, Saskatoon, Canada S7N 0W0.

Successor to Weidenreich

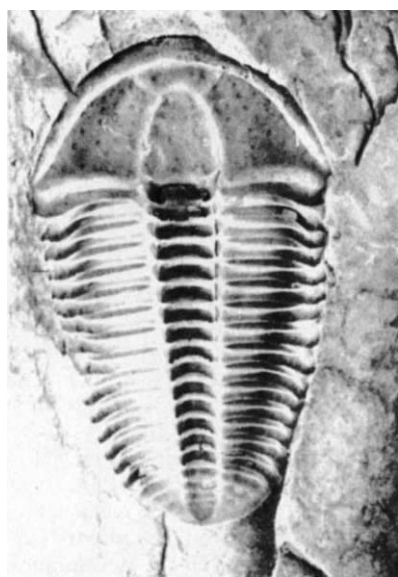
Eric Delson

Koobi Fora Research Project. Volume 4: Hominid Cranial Remains. By Bernard Wood. Oxford University Press: 1991. Pp. 466. £95, \$220.

THE fossil hominids (or more precisely hominins) recovered by Richard Leakey and colleagues from the northeast shore of Lake Turkana have been an important source of data and controversy in human palaeontology since 1969. The Koobi Fora deposits have yielded specimens allocated to the East African 'robust' australopithecine *A. boisei* and tentatively to the generalized *A. afarensis*, as well as to *Homo habilis* and *H. erectus*, whose definitions in fact are a main focus of this work. Although the specimens were usually briefly described after their discovery, it is only here that further details of their morphology and a comparative overview of their relationships are presented. Palaeo-anthropological monographs on human fossils are rare, but in the past two years G. P. Rightmire has perhaps too succinctly reported his worldwide survey of *H. erectus* (*The Evolution of Homo erectus*, Cambridge University Press, 1990), while P. V. Tobias has perhaps too verbosely detailed the fossils from Olduvai Gorge, Tanzania, attributed to *H. habilis* (*Olduvai Gorge IV*, Cambridge University Press, 1991). Wood's treatise is the best of these monographs and is a worthy successor to F. Weidenreich's series of reports on the Zhoukoudian remains (see, for example, *Palaeontol. Sinica, New Ser. D.* 10, 1; 1943), long the standard of comparison in the field.

F. H. Brown and his collaborators have recently surveyed in detail the chronology and palaeoenvironments of Koobi Fora, leaving Wood to concentrate on morphology. After listing all cranial (including dental) fossils by specimen number, Wood reviews previous analyses of early hominin taxonomy with a dozen tables of features distinguishing hominin species from one another. The meat of the book lies in Part II, three chapters providing descriptions and comparisons of the 126 Koobi Fora fossils. This section is organized by body part, but the breakdown makes it slightly hard to use. Wood first discusses the eight well-preserved crania and their attached teeth, followed by less-complete cranial fragments; the mandibular remains are treated similarly. But the isolated teeth are described separately, which may hinder direct comparison. All the metric

Tribute to trilobites



If every picture is worth 1,000 words, *Trilobites* has 120,000 exclusive of text and captions. The author, Harry Whittington, presents us with a beautiful photographic guided tour of the trilobites he has examined during his long and distinguished career. The book is a kind of palaeontological *Desert Island Discs* in which each photograph has been selected to encapsulate a particular scientific point, whether about morphology of a single species or the preservation of fossils in general.

Trilobites were marine, jointed-legged creatures that looked a bit like modern crustaceans. They dominated the sea from their first appearance, back in the Lower Cambrian around 550 million years ago; diversified to occupy habitats from the plankton to the abyssal benthos; and survived until the great extinction that closed the Permian period, some 300 million years later. Longer-lived than the dinosaurs, and far more common as fossils, the trilobites epitomize

the entire Palaeozoic era. Indeed, they represent many people's first experience of fossil collecting.

This is why the appearance of this book should be welcomed. At first glance it has everything going for it — a comprehensive guide by an acknowledged authority, and the second volume in a series of illustrated guides to fossils aimed at a general readership. But apart from the photographs (which speak for themselves), *Trilobites* is a far less friendly proposition than *Graptolites*, edited by D. Palmer and B. Rickards, the book that opened the series (for a review see *Nature* 350, 202; 1991). There are very few concessions to the nonspecialist: too often one finds oneself enmeshed in the finer points of trilobite anatomy, almost always without a map or, often, a reason why. Clearly, Whittington is addressing not a general audience, but his peers.

It is ironic that the book's biggest flaw concerns a picture — or rather, the lack of one. There is no single, simple, labelled diagram in which the anatomy of trilobites is explained. In its absence, the meanings of terms such as pygidium, glabella, opisthoparian, pleura, hypostome, cranidium and doublure remain hard to grasp except by inference; and the frequent discussion of such arcana without adequate explanation makes the book as a whole a tough read (the introduction to anatomy is itself hard to follow).

A further disincentive to the uninitiated is the price, which won't leave enough change out of £40 to sneeze into. As an album of glorious pictures, however, *Trilobites* would be cheap at half as much again, even if one didn't read the text. But it would be a pity not to do so, given the insight the text offers (albeit after extensive excavation). Published by Boydell, £39.95, \$79.

Henry Gee

data, however, are presented in 30 tables in Part IV, with nearly 400 measurements provided, each carefully defined and readily reproducible. The descriptions here are also useful, although Wood generally avoids transforming his data into verbal characters of the form listed and compared in Part I.

Each fossil, no matter how fragmentary, is well illustrated, and most are allocated to one of the recognized species: the majority (53, plus six recovered after 1981 and not detailed here) to *A. boisei*, but also 24 to *Homo* sp. (something like *H. habilis*) and 17 to *H. aff. H. erectus*. Two specimens were recovered from deposits older than 2.5 million years and have previously been allocated to *A. afarensis*, but Wood is more cautious about their affinities.

The interpretation of the 'robust' australopith sample is widely accepted. But since the recovery of the cranium KNM-WT 17000 on the western side of Lake Turkana, there has been argument about whether it and other early specimens are members of the species *A. boisei*: Wood treads lightly but removes the early (and morphologically conservative) group of fossils from this species' hypodigm (the list of fossils allocated to the species) and seems to treat the group as part of a separate lineage if not as a full species. He also comes down firmly but not unequivocally in favour of monophyly (descent from a single ancestral group) of all 'robusts' (rather than an independent derivation of South African *A. robustus* and resulting facial and dental convergence) and the resuscitation of *Paranthropus* as a full genus for these species. In his sparkling foreword, F. Clark Howell applauds the first two decisions and suggests the use of *Paranthropus* as a subgenus of *Australopithecus*.

Even more argument surrounds the interpretation of the earliest species allocated to *Homo*. The general view in the mid-1980s was that an earlier *H. habilis* was succeeded in Africa by *H. erectus*, which later spread into Eurasia and eventually gave rise to *H. sapiens*. Then several workers, including Wood, began to question whether the variation in *H. habilis* was too great for a single species, as well as whether the African members of *H. erectus* are so morphologically conservative that they should be placed in a species separate from the East Asian archetypes of that taxon.

Wood proposes and tests several alternative hypotheses to explain the distribution of 'gracile' specimens at Olduvai and the Turkana basin between about 2.1 and 1.6 million years ago. He first rejects a single-species solution, and eventually suggests the occurrence of a single species (*H. habilis*) at Olduvai,

found also at Turkana alongside a form with larger brains and teeth, termed "*Homo* sp. indet." In a review article in *Nature* (355, 783; 1992), Wood formally applies the name *H. rudolfensis* (Alexeyev, 1986) to this taxon. As Wood notes, however, various authors have proposed alternative hypodigms for the two forms; Tobias in fact subsumed all these specimens into *H. habilis*. My own concern is that one model seems not to have been fully evaluated: this would allow for the presence of both species at Olduvai, with the partial skulls OH 7 (the holotype of *H. habilis*) and OH 16 as possible specimens of the larger form. Moreover, Figure 6.9 in the book reveals that two crania from Koobi Fora, KNM-ER 1470 and 1813, are more distinct than male and female crania of humans, gorillas or chimpanzees; but Wood does not comment on the near conformity of the relationship between the two fossils and that between the sexes of *Pongo pygmaeus*. Nonetheless, the presence of *H. rudolfensis* alongside *H. habilis* at least in the Turkana Basin now seems well established.

Wood's assessment of the later specimens of *Homo* at Koobi Fora is less satisfying. Although he accepts that the sample is closely similar to *H. erectus*, he argues that it is sufficiently different from the Indonesian and Chinese populations for these to be classed as a distinct species. Wood terms this taxon *H. aff. H. erectus* but notes that *H. ergaster* Groves and Mazak, 1975, is probably the valid name. Yet to me the differences between the African and Asian samples are less clear than those between any of the taxa discussed previously; their real distinction rests mainly on the conservative morphology of the earlier African forms as compared to the distinctive features, or autapomorphies, of the generally younger Asian fossils. Rightmire, and A. Turner and A. Chamberlain (*J. Hum. Evol.* 18, 115; 1989) have made the strongest recent case for the global unity of *H. erectus*, and I find their arguments more convincing.

Clearly, palaeoanthropologists are difficult to please and to convince, but this work should go a long way towards appeasing them. Wood's care and desire for usefulness has resulted in a monograph that epitomizes comparative biology: a source of well-illustrated, reproducible data followed by reasoned assessments of serious alternatives. The volume deserves a place on the shelf of every active researcher and advanced student of human palaeontology. □

Eric Delson is in the Department of Anthropology, Lehman College and the Graduate School, City University of New York, Bronx, New York 10468, USA.

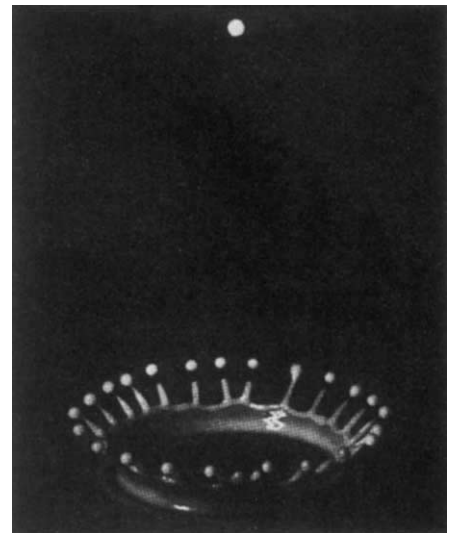
Perturbing patterns

Simon L. Altmann

Fearful Symmetry: Is God a Geometer?

By Ian Stewart and Martin Golubitsky. Blackwell: 1992. Pp. 287. £16.95, \$21.95.

If you have a perfectly circular bowl of milk and let a drop fall exactly in the centre of its circular surface, conservation of symmetry (alas, naïvely used) would predict that in the resulting splash



Splash in the pan — symmetry breaking as illustrated in D'Arcy Thompson's *On Growth and Form*. An abridged edition of the book is newly published in paperback by Cambridge University Press, £7.95.

a cylindrical rim is raised around the circumference of the bowl. A beautiful photograph first published by D'Arcy Thompson shows that, instead, a crown-shaped transient object is created with 24 spikes. Here we have a symmetry paradox, which allows the authors to discuss the use and misuse of the principle of symmetry of Pierre Curie and the problem of symmetry breaking. A short discussion on the relation between symmetry and geometry on the one hand and group theory on the other, helps us to understand how, when symmetry is broken, a descent of symmetry occurs from the original object, which belongs to a given symmetry group, to a deformed state of the object that belongs to one of the subgroups of the original group. All this sounds like heady stuff, but the authors bring out the group concept gently, although not without decorum by means of practical examples. They even manage to delve into the intricacies of crystal symmetries, for which systems they again discuss the book's favourite sub-

ject of symmetry breaking.

Fluid flow has long been a rich source of symmetry paradoxes, as they used to be called before symmetry breaking became both better understood and fashionable. A useful chapter deals with these problems and forms a good introduction to the chapter entitled "The Universe and Everything", where we are reminded of the original sin: "The most spectacular broken symmetry of them all is our own Universe".

One of the most fascinating cases of broken symmetry arises in the problem of the growth of biological structures. Although these systems may start by being fairly homogeneous, this homogeneity becomes unstable, thus allowing patterns and structures to emerge. This is the observation that led Alan Turing to his famous mathematical theory of morphogenesis. But more surprises await us in the way in which symmetry breaking creeps into everything: when a horse stands still it shows, with any luck, bilateral symmetry, that is, there is a symmetry plane cutting along the length of the animal through its centre. When the horse walks, however, this symmetry is destroyed: we are thus entertained with a full chapter on gait with examples from throughout the animal kingdom. Chaos has been one of the developments in mathematics since the Second World War that has engaged more popular interest, and it is not surprising that chaos, bifurcation and symmetry breaking are all intimately connected.

We are now approaching the last chapter of this entertaining book and the authors have still not commented on the geometric proclivities of the deity. This is a great and relevant question, especially now that Roger Penrose has declared himself a Platonist in mathematics. Mathematics is so unreasonably effective, in the words of Eugene Wigner, in describing the Universe, that one is highly tempted to see in it the stamp of the Creator. The authors are too well mannered to try to impose their views on us, but very discreetly, in two lines, they neatly dispose of the problem: Wigner's puzzle, they say, "may have a simple answer: mathematics is effective in describing the universe because that's where we got it from". It is a pity that they do not produce evidence for this statement. Could it be that they are keeping it for their next book?

All in all, the book is a very good read: I would have enjoyed even more a more restricted range of subjects treated a little more sparsely, but those who want a bit of everything will find plenty to attract their curiosity. □

Simon L. Altmann is at Brasenose College, Oxford OX1 4AJ, UK.

Neighbours

M. G. Edmunds

The Andromeda Galaxy. By Paul Hodge. Kluwer: 1992. Pp. 358. DFL 160, \$79, £55.

OUR 'big sister' Andromeda is one of a handful of galaxies that can be seen with the naked eye. Many readers may recall its being pointed out to them when they were young, with hushed tones reminding them that its light had started out on its two-million-year journey before humans had evolved.

The serious study of Andromeda began more than a hundred years ago, when a supernova appeared in the galaxy in August 1885. Until the revolution in light-detection techniques of the past 30 years, detailed observations required heroic efforts — 80-hour exposures for photographic spectra were not unknown. Exploration was dominated by two of the great characters of twentieth-century astronomy, Edwin Hubble and Walter Baade. Since the 1960s, extensive studies have been carried out, particularly by Sidney van den Bergh and Paul Hodge himself, with a vast increase in both data and the number of Andromeda acolytes.

Although it is really its proximity that makes Andromeda special, its family resemblance to our own Galaxy is an obvious attraction. It is perhaps twice as massive and 40 per cent larger than the Milky Way, but appears to be lazier, forming new stars at a considerably slower rate — only about one-tenth of that of our own Galaxy. This weaker star formation is concentrated, for no appa-

rent reason except that interstellar gas is concentrated in a similar way in a broad ring midway out in the disk. A proper understanding of the causes of the different rates of star formation in galaxies remains an important problem in astrophysics.

Is a compilation of observations on a single object really helpful? A biography of a person can sometimes illuminate history much better than a collection of general works, and at least this book forces one to try to interrelate an extraordinarily wide range of galactic phenomena. The book is very much a well-illustrated review and betrays some signs of hasty final preparation. But Hodge writes with a lightness of touch, providing a huge amount of material in a surprisingly readable way. As a source of references the volume is invaluable; Hodge is particularly good in not neglecting the oft-forgotten work from the former Soviet Union. His is an observer's approach, presenting the observations with little — one might say refreshingly little — theoretical interpretation. He confines comparisons to a few members of our own local group of galaxies and intentionally omits discussion of Andromeda's little elliptical companions.

The reader with a moderate background in the astronomy of galaxies is bound to learn (or re-learn) a lot about Andromeda. It is our nearest spiral-galaxy companion, and surely it is a good thing to get to know one's neighbours better. □

M. G. Edmunds is in the Department of Physics and Astronomy, University of Wales College of Cardiff, PO Box 913, Cardiff CF2 3YB, UK.

New references

- **Dictionary of Inorganic Compounds.** Contains physical, structural and bibliographical data on about 42,000 compounds, as well as providing synthetic routes and hazard data. Chapman & Hall, 5 volumes (1–3 entries, 4 & 5 Indexes), 5,400 pages. Pre-publication price, £2,350 (£2,750 for orders after 31 December 1992).
- **Academic Press Dictionary of Science and Technology.** Contains some 124,000 fully defined entries, about 2,000 illustrations, and brief essays by "prestigious scientists" on 124 different fields. Academic, 2,432 pages, £68.
- **Encyclopedia of Marine Sciences** edited by J. G. Baretta-Bekker, E. K. Duursma and B. R. Kuipers. Contains about 3,000 entries. Springer, DM 58 (pbk).
- **Free-Living Freshwater Protozoa: A Colour Guide** by D. J. Patterson and S. Hedley. Wolfe, £48.

New in paperback

- **Deep-Sea Biology: A Natural History of Organisms at the Deep-Sea Floor** by J. D. Gage and P. A. Tyler. Cambridge University Press, £24.95, \$39.95. For a review see *Nature* **352**, 294 (1991).
- **Dinosaur Systematics: Approaches and Perspectives** edited by K. Carpenter and P. J. Currie. Cambridge University Press, £17.95, \$29.95.
- **Infectious Diseases of Humans: Dynamics and Control** by R. M. Anderson and R. M. May. Oxford University Press, £22.50, \$47.50. For a review see *Nature* **358**, 29 (1992).
- **The Creative Mind** by M. A. Boden. Basic Books, \$15. For a review see *Nature* **349**, 378 (1991).
- **Our Universe: An Armchair Guide** by Michael Rowan Robinson. Freeman, \$19.95. For a review see *Nature* **348**, 355 (1990).
- **Centuries of Darkness** by P. James. Pimlico, £12.99. For a review see *Nature* **353**, 712 (1991).

Single-electron transfer in metallic nanostructures

Michel H. Devoret, Daniel Esteve & Cristian Urbina

Electrons can be made to pass through a circuit one by one, in nanoscale devices based on the combination of the Coulomb interaction between electrons and their passage by quantum tunnelling through an insulating barrier. Single-electron devices provide a new way of measuring the charge quantum, and clarify how electronic signal processing at the molecular level might function.

ONE individual atom, a purely theoretical entity a hundred years ago, can now be imaged and manipulated at the surface of bulk matter¹ or, free-standing, in vacuum². Is the electron, the simplest and most thoroughly studied particle, amenable to such ultimate control? In vacuum, the detection of single electrons is now routine. A spectacular example of the control of individual electrons travelling in a vacuum chamber is the experiment in which Dehmelt *et al.*³ were able to probe during three months a single electron kept in an electromagnetic trap, thereby measuring to unprecedented accuracy the anomalous part of its magnetic moment. In matter, the manipulation of individual electrons is a very different game, because the separation between electrons is of the same order as their quantum mechanical wavelength. Here we focus on the most basic type of such manipulation. We explain how it is possible to take, at a precise instant, exactly one electron from a first electrode and transfer it with certainty to a second electrode. By making these electrodes part of an electrical circuit and by continuously repeating this transfer process we can achieve a perfectly controlled current source. In particular, for a sequence of single-electron transfers clocked by a radiofrequency signal at frequency f , the current I will be given simply by $I = ef$ where e is the quantum of charge, a fundamental constant.

Basics of single-charge transfer

Although the charge of the electron was measured as early as 1911 (ref. 4), the granularity of electricity does not usually show up in the macroscopic quantities such as current and voltage, which describe the state of an electric circuit. This is not just a matter of the number of electrons being very large in typical devices. Charge flow in a metal or a semiconductor is a continuous process because conduction electrons are not localized at specific positions. They form a quantum fluid which can be shifted by an arbitrarily small amount. The variations of the charge Q on a capacitor C and of the associated potential difference $U = Q/C$ illustrate this property. The charge Q can be any fraction ε of the charge quantum e : if ρ denotes the electron density in the metallic plates of the capacitor and S their surface area, it is easy to see that a bodily displacement $\delta = \varepsilon/(\rho S)$ of the electronic fluid with respect to the ionic background, in the direction perpendicular to the plates, produces the charge $Q = e\varepsilon$.

There exists, however, a solid-state device in which electric charge flows in a discrete manner. It consists of two metallic electrodes separated by an insulating layer so thin that electrons can traverse it by the tunnel effect⁵ (Fig. 1). Tunnelling can be considered as an all-or-nothing process because electrons spend a negligible amount of time under the potential barrier corresponding to the insulating layer^{6,7}. If one applies a voltage V to such a tunnel junction, electrons will randomly tunnel across the insulator at a rate given by V/eR_t , where the tunnel resistance R_t is a macroscopic parameter of the junction which depends on the area and thickness of the insulating barrier. Apart from allowing the tunnel effect, the two facing electrodes

behave as a capacitor whose capacitance C is the other macroscopic parameter of the junction. It is important to stress that the transport of electrons in a tunnel junction and in a metallic resistor are fundamentally different, even though the current-voltage characteristic is linear in both cases. Charge flows continuously along the resistor, whereas it flows across the junction in packets of e . Obviously, a tunnel junction provides the means to extract electrons one at a time from an electrode. With a single voltage-biased tunnel junction, however, it is not possible to control the instants at which electrons pass from the upstream electrode to the downstream electrode, because of the stochastic nature of tunnelling. A further ingredient is needed.

Suppose that instead of applying directly a voltage source to the junction one biases it with a voltage source U in series with a capacitor C_s (we reserve the letter symbol V for transport voltage sources that have to deliver a static current). A metallic electrode entirely surrounded by insulating material is formed between the junction and the capacitor (Fig. 2a). We will call such an isolated electrode, which electrons can enter and leave

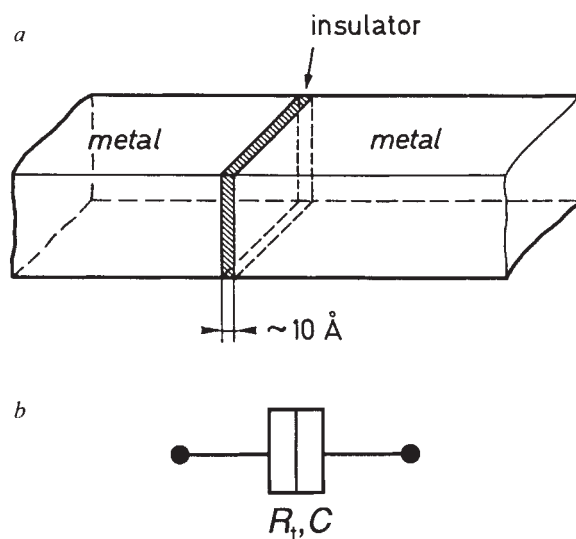


FIG. 1 a, Tunnel junction traversed by a current $I(t)$ which consists, when a fixed voltage is imposed to the junction, of uncorrelated charge packets corresponding to individual electrons. Electrons tunnel through the thin layer of insulator sandwiched between the metal electrodes. The junction is represented in circuit schematics by a double box symbol (b) and is characterized by the tunnel resistance R_t and capacitance C . It is worth noting that although R_t is called a 'resistance', it characterizes a purely elastic process. At the insulating barrier, the electron wavefunction is partially transmitted and reflected. Its energy does not change. The tunnel resistance is inversely proportional to the barrier transmission coefficient which decreases exponentially with the thickness of the insulating layer. In practice, measurable tunnel resistances can be achieved only with insulating layers a few nanometres thick.

only by tunnelling, an 'island'. The island is coupled electrostatically to the rest of the circuit by the capacitances C and C_s whose charges are denoted by Q and Q_s respectively. Although, as we have seen, Q and Q_s are both continuous variables, their difference is the total excess charge of the island. Because charge can enter the island only by tunnelling, this total charge is a multiple of the electron charge: $Q - Q_s = ne$. Suppose furthermore that the island dimensions are small enough that the electrostatic energy $E_c = e^2/2C_\Sigma$ of one excess electron on the island is much larger than the characteristic energy $k_B T$ of thermal fluctuations. Here, $C_\Sigma = C + C_s$, and k_B and T denote the total capacitance of the island, the Boltzmann constant and the temperature, respectively. This Coulomb energy E_c is the other ingredient of controlled electron transfer.

When $U = 0$, n will stay identically zero because the entrance or exit of an electron would raise the electrostatic energy of the island to a level much higher than permitted by thermal fluctuations. As U increases from zero, however, the total energy difference between the $n = 0$ and $n = 1$ state of the whole circuit decreases, because when an electron tunnels to the island the potential drop $C_s U / C_\Sigma$ partly compensates the electrostatic energy of the island. In fact, a straightforward calculation of the total energy of the circuit yields $E = E_c(n - C_s U / e)^2$. Thus, when $U = e/2C_s$, the $n = 0$ and $n = 1$ states will have the same energy and an electron can tunnel in and out freely. As U is increased further, the $n = 1$ state becomes the lowest energy state. The maximum stability of the $n = 1$ state against fluctuations is reached at $U = e/C_s$ where, as in the case $U = 0$, and $n = 0$, the charge Q vanishes. It is now easy to see that each time the voltage U is increased by e/C_s , the number n of excess electrons of the island is increased by one. If one plots $\bar{n}$, the average of n , as a function of U , one gets the staircase function shown in Fig. 2b. It is therefore possible to control exactly the number of excess electrons of the island by adjusting the voltage U .

As the temperature is increased, the staircase becomes rounded and for temperatures $k_B T \gg E_c$ it approaches the straight dotted line of Fig. 2b. In practise, one can reliably cool tunnel junctions down to 50 mK but not much below. To satisfy $E_c \gg k_B T$, C_Σ must be of the order of or smaller than one femtofarad. This requires the fabrication of junctions with typical areas of $50 \text{ nm} \times 50 \text{ nm}$ and hence the use of nanofabrication techniques. With such low values of capacitance, the typical voltage corresponding to the addition of an electron is of the order of $100 \mu\text{V}$ – 1 mV , a value which can be easily controlled electronically. To summarize, tunnelling breaks the continuity of the electron fluid into charge packets corresponding to single electrons. The Coulomb energy of excess charges on an island provides a feedback mechanism that regulates the number of electrons tunnelling in and out of the island. At sufficiently low temperature, the exact number of excess electrons on the island does not fluctuate and can be entirely determined by an externally applied voltage. The quenching of the island charge fluctuations for the 'single electron box' (the circuit of Fig. 2a) has been demonstrated experimentally by Lafarge *et al.*⁸.

We have considered so far only thermal fluctuations of the number n . This variable is also subject to quantum fluctuations. In our analysis of the circuit of Fig. 2a, we have neglected the delocalization energy associated with tunnelling. This energy is very small compared with the Coulomb energy. Perturbative calculations^{9,10} show that the quantum fluctuations of n become negligible in the limit $R_t \gg R_K = h/e^2$, where h is Planck's constant. The constant $R_K \approx 26 \text{ k}\Omega$ is the resistance quantum. In the next three sections we will consider tunnel barriers sufficiently opaque that this latter condition is fulfilled.

Single electron effects: a brief history

A large class of phenomena exist that combine the partial localization of electrons due to tunnelling and the Coulomb charging energy, and may be called 'single-electron effects'.

Decades ago, it was proposed that the variation of the island potential due to the presence of only one excess electron could be large enough to react back on the probability of subsequent tunnelling events^{11–15}. At that time, the effect could only be observed in granular metallic materials. It was realized that the hopping of electrons from grain to grain could be inhibited at small voltages if the electrostatic energy of a single electron on a grain was much larger than the energy of thermal fluctuations. The interpretation of these pioneering experiments, in which there is an interplay between single-electron effects and random media properties, was complicated by the limited control over the structure of the sample. With modern nanofabrication techniques, it is possible to design metallic islands of known geometry separated by well-controlled tunnel barriers¹⁶. This led Fulton and Dolan to perform the first unambiguous demonstration of single-electron effects in an island formed by two junctions¹⁷. Meanwhile, Likharev and coworkers^{18,19} had produced detailed predictions of single-electron effects in a nanoscale current-biased single junction (this system was also considered in refs 20 and 21, but only for junctions in the superconducting state) and proposed various applications of the new effects. This current-biased scheme is analogous in some ways to the circuit of Fig. 2a, but with the capacitor replaced by a resistor. In that case there is no island enforcing charge quantization, because an arbitrarily small amount of charge can flow through the resistor. Only the charge on the junction capacitor would provide the feedback of Coulomb energy on tunnelling.

It was later understood that the quantum electromagnetic fluctuations due to the resistor wash out single-electron effects

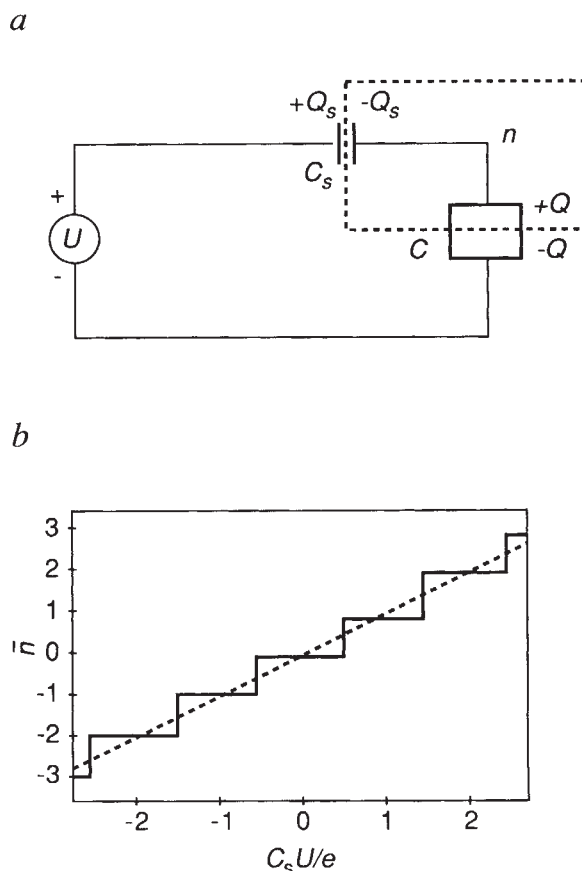


FIG. 2 a, Junction biased by a voltage source U in series with a capacitance C_s . The metal electrode between the junction and the capacitance forms an isolated 'island' (box in dashed line) which contains n excess electrons. b, Variation of $\bar{n}$, the average of n as a function of U when $k_B T \ll E_c$ (full line) and $k_B T \gg E_c$ (dashed line).

in this single-junction no-island system, unless the resistor is much larger than the resistance quantum R_K up to frequencies of the order of $e^2/(hC)$ (refs 22–24). This has been partially achieved and the competition between single-electron effects and quantum electromagnetic fluctuations has been observed^{25,26}. The single-junction no-island system is of interest as an illustration of the foundations of the field, but it is not suited for practical applications because getting rid of quantum electromagnetic fluctuations is so difficult experimentally. In what follows, we will resume the discussion of systems that contain at least one island and are thus immune to quantum electromagnetic fluctuations. We will focus mainly on the controlled transfer of single electrons. For general introductions to single-electron effects in normal and superconducting junction systems, see refs 28–30; for recent snapshots of the state of current research, see refs 31, 32.

The single-electron transistor

The one-junction one-island circuit of Fig. 2a is the simplest in which single-electron transfer can occur. On the other hand, it cannot produce an externally measurable static current, as the island is a cul-de-sac for electrons. Let us consider the next order of complexity, the two-junction one-island circuit of Fig. 3a (ref. 17). The state of the circuit is now characterized by the two numbers N and N' of electrons having passed through the two junctions. (The sign of N and N' is positive if during tunnelling the electron flows in the direction of increasing voltage, and negative otherwise.) It is convenient to introduce the number $n = N - N'$ of excess electrons on the island and the charge flow index $p = (N + N')/2$. The state $(n, p + 1)$ only differs from the state (n, p) in that one electron has been transferred from one terminal of the transport voltage source V to the other. The electrostatic energies of the various capacitances of the circuit are the same. As the precise value of p does not matter here, we will condense the notations (n, p) and $(n, p + 1)$ into (n) and $(n)^*$. With regard to the total energy of the circuit, which includes the work of the transport voltage, state $(n)^*$ is lower by eV than state (n) and, hence, the circuit has no absolutely stable states. In principle, a steady current I could flow around the loop formed by the two junctions and the transport voltage V . To go from state (n) to state $(n)^*$, however, the circuit must go through state $(n + 1)$ or state $(n - 1)$, because tunnel events occur one at a time. States $(n + 1)$ and $(n - 1)$ differ from state (n) by an electron having tunneled through the first and second junction, respectively. This is where the single-electron Coulomb energy $E_c = e^2/2C_\Sigma$ comes into play (C_Σ is, as before, the island total capacitance given now by $C_\Sigma = C + C' + C_g$, where C and C' are the two junction capacitances). To simplify the discussion, suppose that $eV \ll E_c$.

When the control 'gate' voltage is set at $U = 0$, the energy of states (-1) and (1) will be $E_c - eV/2 \approx E_c$ above the energy of state (0) (see Fig. 3c). At low temperature, this will provide a Coulomb barrier for the transport of electrons around the circuit. In this case the current I should be strictly zero. This situation is called the Coulomb blockade. On the other hand, when the control voltage is such that $C_g U \approx e/2$, states (0) and (1) have nearly the same energy (Fig. 3c). As soon as the energy of the (1) state is lowered below that of the (0) state, the $(0) \rightarrow (1)$ transition becomes possible and an electron enters the island through the first junction. If U is such that the energy of the (1) state, although below that of the (0) state, is still above the energy of the $(0)^*$ state, the transition $(1) \rightarrow (0)^*$ takes place and the electron leaves the island through the second junction. Apart from an electron having gone through the device, one is now back to the initial electrostatic state and the cycle can start over again. This cascade of transitions produces a current of order $V/(R_1 + R_2)$ through the device (R_1 and R_2 are the tunnel resistances of the two junctions). When U is increased further, the energy of the (1) state goes below the energy of the $(0)^*$ state and one enters a new Coulomb blocked state with one excess

electron on the island. The domains of the Coulomb blocked states in the set of U values are in a one-to-one correspondence with the flat portions of the staircase of the electron box (Fig. 2b) and it is easy to show that, at voltages low compared with the Coulomb voltage E_c/e , the current I is maximum when $C_g U = ne/2$.

In practise, a current of the order of 10^9 electrons per second can be switched on and off by the presence or absence of half the electron charge on the gate capacitor, hence the name 'single electron transistor' (SET) given to this device. The remarkable charge sensitivity of the SET is unrivalled by other devices. (The charge sensitivity of the SET electrometer is six orders of magnitude better than conventional FET electrometers²⁸. A possible application is the detection of individual photoinduced electron-hole pairs in semiconductors³³.) But the input capacitance of the SET is, by construction, so tiny that its voltage sensitivity is not high. In this respect, it does not compare favourably with the field effect transistor (FET), the semiconductor device on which most of today's applications of solid-state electronics are based. Furthermore, in the SET the modulation of electron flow by the gate ceases as soon as the bias voltage becomes of the order of the Coulomb gap voltage E_c/e , whereas in the FETs used in digital circuits the modulation of the source-drain current by the gate only saturates at large bias voltages³⁴. It is this latter feature which ensures enough voltage gain to compensate

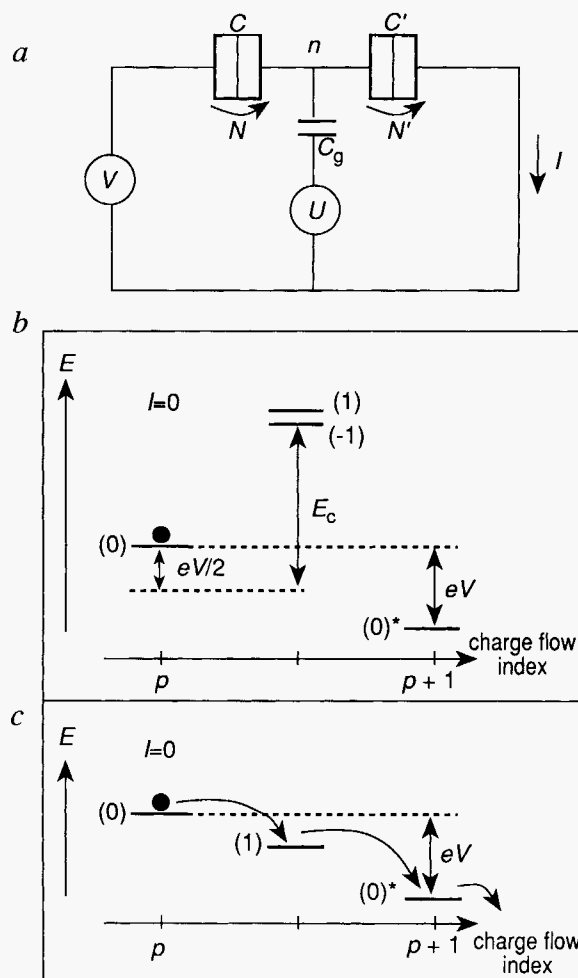


FIG. 3 a Schematic of single-electron transistor (SET). Energy of the states of the circuit when (b) $U=0$ and (c) $U=e/C_g$. The numbers in parenthesis are the values of the number n of excess electrons on the SET island. The charge flow index is half the sum of the numbers N and N' of electrons that have traversed the junctions. In (b) no current can flow through the device: this is the Coulomb blockade.

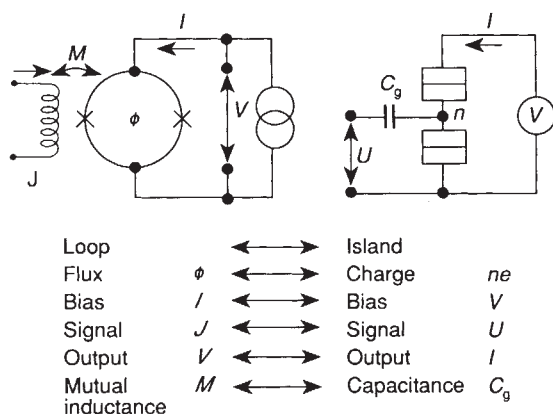


FIG. 4 Comparison between the d.c. SQUID (left) and the SET electrometer (right).

for the dispersion in device parameters and which make robust integrated digital circuit design possible with FETs.

An analogy²⁸ can be drawn between the SET and the d.c. SQUID with an input coil³⁵ (see Fig. 4). The d.c. SQUID (superconducting quantum interference device) consists of two Josephson junctions in parallel biased by a static current. In the d.c. SQUID, the output voltage is a periodic function of the current in the input coil, whereas in the SET, the output current is a periodic function of the voltage on the input capacitor. For the d.c. SQUID the period is set by the flux quantum $h/2e$ whereas for the SET the period is set by the charge quantum e . It is tempting to speculate that the SET will play the same role for ultra-sensitive electrometry that the d.c. SQUID plays for ultra-sensitive magnetometry. However, the fundamental impossibility of building the charge analogue of the superconducting flux transformer which is so crucial to the use of d.c. SQUIDS may severely limit the use of SETs.

The junctions that have been described so far consist in practice of two overlapping metallic films. It is also possible, instead of the three-dimensional gases that conduction electrons form in a metal, to use two-dimensional gases which are found in semiconductor heterostructures such as GaAs/GaAlAs. The detailed manifestations of Coulomb blockade have been thoroughly studied in these systems, where single-electron effects may coexist with the quantum Hall effect³⁶.

Finally, Coulomb blockade has been observed with a scanning tunnelling microscope (STM) placed over a tiny metallic drop-

let³⁷. The role of the island is played by the droplet. Unfortunately, it has so far been impossible to modulate the gate voltage independently in the droplet-STM systems. On the other hand, very small island dimensions (a few nanometres) can be achieved in this manner, and Coulomb blockade at room temperature has been reported³⁸. In principle the island could even be reduced to a single molecule.³⁹.

Controlled transfer of charges in a circuit

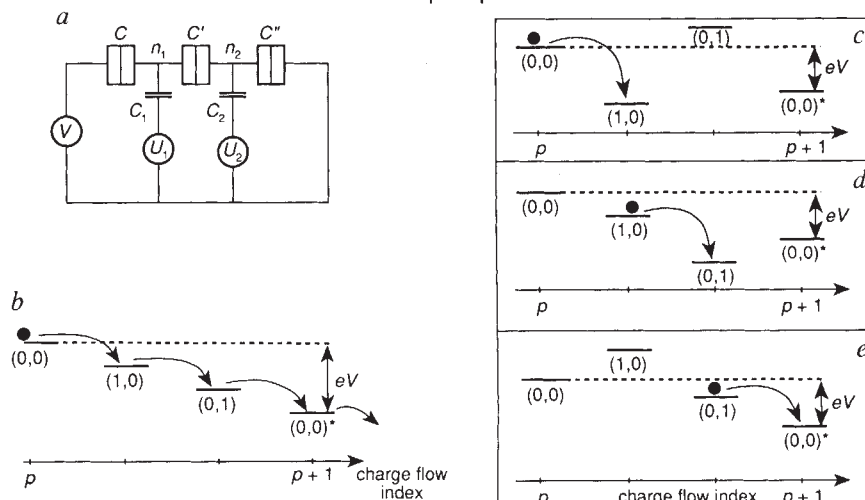
Although the principle of the SET involves the electrostatic energy of a single electron on the SET island, the charge flow through this device is not controlled at the single-electron level. The voltage U controls only the average value of the current. The instants at which electrons pass through the device are random, as in a single junction. A control of the charge flow electron by electron would mean that, using the control voltage U , one would make a single electron enter the island from the left junction, hold it in the island for an arbitrary time and finally make it leave the island through the right junction. One could then go continuously from a Coulomb-blocked state with $n = 0$ to a Coulomb-blocked state with $n = 1$. This is not possible with only one island. When the energy of the (1) state dips below the energy of the $(0)^*$ state to which it can decay (see Fig. 3c). An electron cannot be made to enter the island through one junction without setting the electrostatic energies so that it is energetically favourable for another electron to leave the island through the other junction.

The control of charge flow at the single-electron level requires at least three junctions^{9,40}. Let us consider the three-junction two-island circuit of Fig. 5a. As in the case of the SET, the state of the circuit can be described using the numbers n_1 and n_2 of excess electrons on each island and the charge flow index given by the third of the algebraic sum of the number of electrons having tunneled through each junction. Using the condensed notation defined above, (n_1, n_2) and $(n_1, n_2)^*$ denote two states whose charge flow indices differ by one, that is, states differing by an electron which has lost energy eV by passing through the entire device.

We suppose $V \ll \min(e/C_{\Sigma 1}, e/C_{\Sigma 2})$ where $C_{\Sigma 1}$ and $C_{\Sigma 2}$ denote the total capacitances of the two islands. The two control voltages U_1 and U_2 , applied to the two gate capacitances C_1 and C_2 , allow us to change the relative energy of the various states of this circuit. If we set U_1 and U_2 to $e/2C_1$ and $e/2C_2$ respectively, the energies of states $(0, 0)$, $(1, 0)$, $(0, 1)$ and $(0, 0)^*$ form a cascade (Fig. 5b). We are in a situation equivalent to the suppression of Coulomb blockade depicted in Fig. 3c, and

Electron pump

FIG. 5 a, Schematic of the single-electron pump. b, Energy states of the circuit when the control voltages U_1 and U_2 are set so that Coulomb blockade is suppressed. c-e, Pumping cycle which transfers one electron around the circuit of a. It is obtained by superposing two phase-shifted modulation signals on the values of U_1 and U_2 corresponding to b.



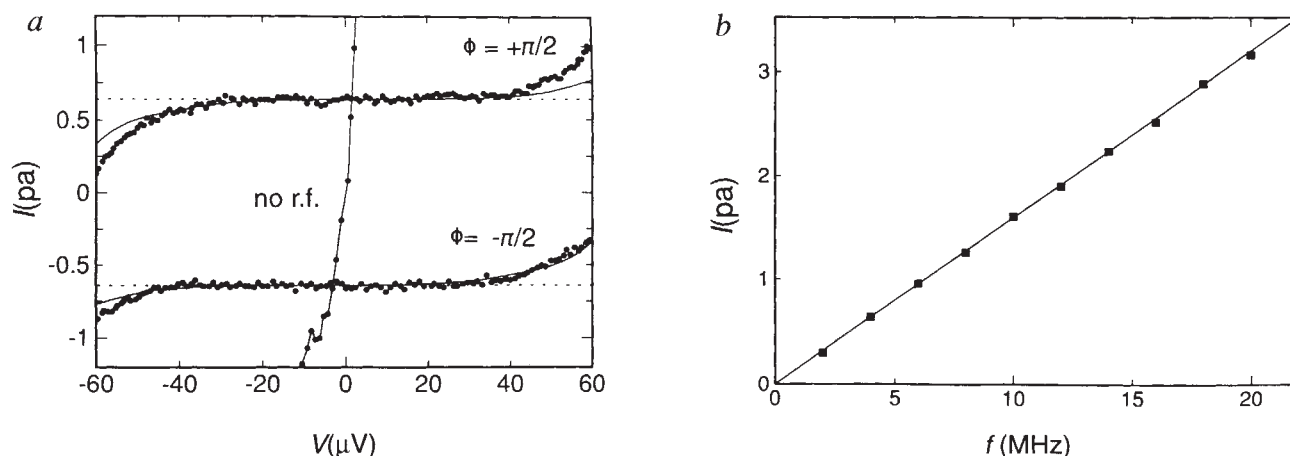


FIG. 6 *a*, Current-voltage characteristic of the pump with and without a $f=4$ MHz control voltage modulation. The two modulation signals were phase-shifted by Φ . Dashed lines indicate $I = \pm ef$. Full lines are the result

of numerical simulations taking into account quantum fluctuations of the island electron number. *b*, Current measured at the inflexion point of the current plateau as a function of the frequency f . Full line is $I = ef$.

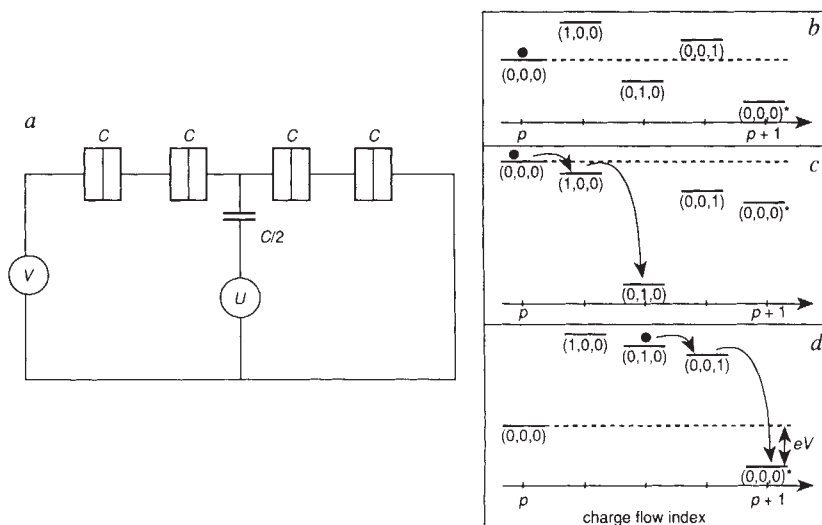
a stochastic current flows through the device. Because there are three junctions instead of two, there are now two intermediate states $(0, 1)$ and $(1, 0)$ in the cascade. Each of these states is coupled to $(0, 0)$ or to $(0, 0)^*$ but not to both. The lowering of either $(0, 1)$ or $(1, 0)$ below $(0, 0)$ and $(0, 0)^*$ stops the stochastic current and puts the circuit in a blocked state. By modulating U_1 and U_2 with dephased periodic signals, the energy of these intermediate states can be cyclically lowered below that of the $(0, 0)$ and $(0, 0)^*$ states while avoiding the cascade configuration of Fig. 5*b* (Fig. 5*c-e*). One starts from the situation where both $(1, 0)$ and $(0, 1)$ are above $(0, 0)$ and $(0, 0)^*$. The circuit is in a blocked state with no excess electrons on the islands. At first, an increase of U_1 lowers $(1, 0)$ below $(0, 0)$ and $(0, 1)$. An electron goes through the left-most junction and the circuit adopts a new blocked state with an extra electron on the first island (Fig. 5*c*). Then U_2 increases while U_1 decreases: this lowers $(0, 1)$ below $(1, 0)$ and $(0, 0)^*$. A tunnel event consequently takes place through the middle junction and the circuit now adopts a blocked state with an extra electron on the second island (Fig. 5*d*). Finally U_2 is decreased to its initial value, making $(0, 1)$ pass above $(0, 0)^*$. An electron goes through the right-most junction and, apart for a charge e having crossed the entire device, the circuit returns to its initial blocked state (Fig.

5*e*). If the transport voltage V is reversed, the same modulation cycle will still carry electrons in the same direction, provided that the energy difference eV between $(0, 0)$ and $(0, 0)^*$ stays small compared with the energy excursions of $(0, 1)$ and $(1, 0)$. The charge now flows in a direction opposite to that imposed by V . Energy conservation is of course not violated. The work done to 'charge' the transport voltage source is provided by the control voltage sources. We have therefore nicknamed this three-junction device the single-electron 'pump'. The pump is reversible: a time-reversed modulation cycle will transfer electrons from right to left.

The actual operation of a physical device is shown in Fig. 6. We first set U_1 and U_2 to the static values $U_1^{dc} = e/C_1$ and $U_2^{dc} = e/C_2$ corresponding to a maximum zero-voltage conductance (centre curve, marked 'no r.f.'). Two periodic signals with the same frequency f but dephased by $\Phi \approx \pi/2$ are then superimposed on the static components U_1^{dc} and U_2^{dc} . This implements the cycle shown in Fig. 5*c-e* and a current plateau is observed (see Fig. 6*a*). One can easily reverse the cycle, leaving all other conditions the same, by changing Φ to $\Phi + \pi$. A current plateau is again observed, with the same absolute value at $V=0$ but with opposite sign. The height of the plateau is plotted against frequency on Fig. 6*b*. The relation $I = ef$ is well verified, further

Electron turnstile

FIG. 7 *a*, Schematic of single-electron turnstile. *b-d*, Turnstile cycle which is obtained by modulating the control voltage U and which transfers one electron around the circuit of *a*.



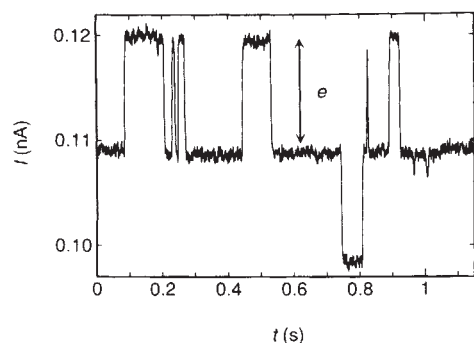


FIG. 8 Time variations of the current through a SET electrometer measuring the charge on an island linked to a charge reservoir through a series of four tunnel junctions. Each jump corresponds to an electron tunnelling into or out of the island.

confirmation that our device does indeed implement the pump principle.

We have seen how two control voltages can transfer electrons one by one in a three-junction device. The transfer of single electron using only one control voltage is possible, but needs at least four junctions. In Fig. 7a we show the schematics of a four-junction three-island circuit which we have nicknamed the 'turnstile'⁴¹. A gate capacitance, with roughly half the value of the capacitance of the junctions, is connected to the central island. Because the gate voltages of the side islands have only to be set to a constant value of zero (in practice, the external gate voltage must be adjusted to compensate for random offset charges²⁸), no gate lines have been represented in the figure. The turnstile can be described as a SET with two junctions in the entrance and exit channels. The intermediate islands create energy barriers whose effect is to suppress the stochastic conduction that takes place in the SET for small V , and for U such that $C_g U = e/2$ (see Fig. 7b–d). For these conditions, the circuit can exist in two states characterized by the presence or absence of an extra electron on the central island. Suppose one starts with no electron in the central island. As U is increased, all the energies of the intermediate states decrease, although the state with an electron on the central island remains the lowest of the intermediate states (Fig. 7b). Consequently, one electron enters the central island. If now one decreases U , all the energies of the intermediate states increase and at one point the state with an extra electron on the central island is no longer the lowest of the intermediate states. An electron then leaves the central island (Fig. 7d). It is easy to see that after one cycle of modulation of U , a charge of one electron has passed through the whole device. Like the pump, the turnstile produces a current $I = ef$, where f is the modulation frequency. Unlike the pump, however, the turnstile is an irreversible device, the sign of the current being imposed by the sign of the bias voltage V .

Metrological applications

We have seen that the pump and the turnstile can produce a current determined only by the frequency f and the quantum of charge e . Because frequencies can be accurately determined, these devices would provide in principle a standard of current. The standard is obtained at present by the combination of the Josephson effect³⁵, which relates a frequency to a voltage through the flux quantum $\Phi_0 = h/2e$, and the quantum Hall effect discovered by von Klitzing⁴², which relates current to voltage through the resistance quantum $R_K = h/e^2$. It is important for metrologists to check whether a direct definition of the ampere using the charge quantum e provided by single electron devices would be compatible with the 'Josephson/Klitzing' definition which combines Φ_0 and R_K . The value of the fine-structure constant $\alpha = e^2/(2\hbar\epsilon_0 c)$, where c and ϵ_0 denote the speed of light and the electrical permittivity in vacuum, is

another important metrological issue that would benefit from the new access to the charge quantum provided by single-electron devices⁴³. This latter application would not require direct measurement of the very low current produced by single-electron devices; one would simply charge a calibrated capacitor with a known number of electrons, and compare its voltage with the Josephson volt.

Although the experiments carried out so far to test the precision of the pump and the turnstile are chiefly limited by the precision of current measurements, we should investigate the intrinsic limitations of the devices. One problem is to ensure that the devices are sufficiently cold while passing current. In that respect the pump principle is better than the turnstile, as the pump is reversible and can operate at zero bias voltage⁹. Theoretical analyses show that the fundamental limitation on the accuracy of the devices is due to co-tunnelling events^{30,44} during which several tunnel events take place simultaneously on different junctions. These higher-order processes are a manifestation of the quantum fluctuations of island electron number discussed above. Fortunately, it can be demonstrated that the rate of co-tunnelling events decreases exponentially with the number of junctions in a device. Several groups have come to the conclusion that an accuracy better than 10^{-8} in the number of transferred electrons is achievable with a pump with five junctions⁴⁵ operating at temperatures of 100 mK or less^{46,47}. An important step towards the practical realization of high-accuracy transfer devices is to show experimentally that the number of electrons on an island is well determined when this island is connected to a charge reservoir through several junctions that block the quantum fluctuations of electron number. We have made a direct measurement of the charge of such an island by using a SET electrometer⁴⁸. In Fig. 8 we show single tunnelling events in and out the island occurring on a timescale of a tenth of a second. Although this timescale is still shorter than expected theoretically, we believe that if smaller junctions were used, the spontaneous tunnel rate could be lowered by two orders of magnitude and thus permit metrological experiments.

Future prospects

It has been suggested^{27,28} that single-electron devices might find applications in digital electronics. A single electron would code for one bit, obviously the most economical way to store information. In fact, the electron pump is already very similar to the shift registers found in computers. The SET would be the building block of this 'single electronics'. A problem, however, is that metallic SETs made using today's technology have no 'engineer gain': one transistor can barely feed one other transistor in the chain of signal processing, once the dispersions on parameters are accounted for. And no one understands how to get rid of random offset charges²⁸ which at present ruin any attempt to have more than a few transistors on one chip. In semiconductor devices, single-electron effects may even appear to be a nuisance because they imply that the electrons go through the dots or channels one at a time, a slowing down of the conventional FET operation. The main benefit of understanding single-electron effects in semiconductor nanotechnology may be just to provide the knowledge to fight them efficiently.

The real virtue of single-electron devices, as far as industrial applications are concerned, is that they teach us how to produce digital functions using only tunnelling and the Coulomb interaction, basic ingredients that are available down to the molecular level. In the as yet undeveloped 'molecular electronics' technology⁴⁹, basic time constants are very short, there is no dispersion in the parameters of individual components and there are few electrons to work with anyway. There, the principles underlying the devices that have been discussed in this article may be fruitfully implemented. The understanding of single-electron effects may stimulate new research directions in the field of

molecular electronics. And after all, was not the search for the 'philosopher's stone', a premature scientific concept in the Middle Ages, the root of modern chemistry? □

Michel H. Devoret, Daniel Esteve and Cristian Urbina are at the Service de Physique de l'Etat Condensé, CEA-Saclay, 91191 Gif-sur-Yvette cedex, France.

1. Eigler, D. M. & Schweizer, E. K. *Nature* **344**, 524 (1990).
2. Wineland, D. J., Itano, W. M. & Van Dyck, R. S. Jr *Adv. atom. molec. Phys.* **19**, 135 (1983).
3. Van Dyck, R. S. Jr, Schwinberg, P. B. & Dehmelt, H. G. *Phys. Rev. D* **34**, 722 (1986).
4. Millikan, R. A. *Phys. Rev.* **32**, 349 (1911).
5. Solymar, L. *Superconductive Tunneling*, Ch. 2 (Chapman and Hall, London, 1972).
6. Büttiker, M. & Landauer, R. *Phys. Rev. Lett.* **49**, 1739 (1982).
7. Persson, B. N. J. & Baratoff, A. *Phys. Rev. B* **38**, 9616 (1988).
8. Lafarge, P. et al. *Z. Phys. B* **85**, 327 (1981).
9. Esteve, D. *Single Charge Tunneling*, Ch. 3 (ed. Grabert, H. & Devoret, M. H.) (Plenum, New York, 1992).
10. Matveev, K. A. *Zh. eksp. teor. Fiz.* **99**, 1598 (1991); (Engl. transl.) *Sov. Phys. JETP* **72**, 892 (1991).
11. Gorter, C. J. *Physica* **17**, 777 (1951).
12. Neugebauer, C. A. & Webb, M. B. *J. appl. Phys.* **33**, 74 (1962).
13. Giaver, I. & Zeller, H. R. *Phys. Rev. Lett.* **20**, 1504 (1968).
14. Lambe, J. & Jaklevic, R. C. *Phys. Rev. Lett.* **22**, 1371 (1969).
15. Kulik, I. O. & Shekter, R. I. *Zh. eksp. teor. Fiz.* **68**, 623 (1975); (Engl. transl.) *Sov. Phys. JETP* **41**, 308 (1975).
16. Dolan, G. J. & Dunsmuir, J. H. *Physica B* **152**, 7 (1988).
17. Fulton, T. A. & Dolan, G. J. *Phys. Rev. Lett.* **59**, 109 (1987).
18. Likharev, K. K. & Zorin, A. B. *J. low Temp. Phys.* **59**, 347 (1985).
19. Averin, D. V. & Likharev, K. K. *J. low Temp. Phys.* **62**, 345 (1986).
20. Widom, A., Megaloudis, G., Clark, T. D., Prance, H. & Prance, R. J. *J. Phys. A* **15**, 3877 (1982).
21. Ben-Jacob, E. & Gefen, Y. *Phys. Lett. A* **108**, 289 (1985).
22. Nazarov, Yu. V. *Pis'ma Zh. eksp. teor. Fiz.* **49**, 105 (1989); (Engl. transl.) *JETP Lett.* **49**, 126 (1990).
23. Devoret, M. H. et al. *Phys. Rev. Lett.* **64**, 1824 (1990).
24. Girvin, S. M., Glazman, L. I., Jonson, M., Penn, D. R. & Stiles, M. D. *Phys. Rev. Lett.* **64**, 3318 (1990).
25. Cleland, A. N., Schmidt, J. M. & Clarke, J. *Phys. Rev. Lett.* **64**, 1565 (1990).
26. Kuzmin, L. S., Nazarov, Yu. V., Haviland, D. B., Delsing, P. & Claeson, T. *Phys. Rev. Lett.* **67**, 1161 (1991).
27. Likharev, K. K. *IBM J. Res. Dev.* **32**, 144 (1988).
28. Averin, D. V. & Likharev, K. K., in *Quantum Effects in Small Disordered Systems* (eds Altshuler, B. L., Lee, P. A. & Webb, R. A.) (Elsevier, Amsterdam, 1991).
29. Schön, G. & Zaikin, A. D. *Phys. Rep.* **198**, 237 (1990).
30. *Single Charge Tunneling* (eds Grabert, H. & Devoret, M. H.) (Plenum, New York, 1992).
31. *Single Charge Tunneling* spec. Issue *Z. Phys. B* **85**, 317–468 (1991).
32. *Single Electron Tunneling and Mesoscopic Devices, Proc. 4th Int. Conf. SQUID '91* (eds Koch, H. & Lübbig, H.) (Springer, Berlin, 1992).
33. Cleland, A. N., Esteve, D., Urbina, C. & Devoret, M. H. *Appl. Phys. Lett.* (in the press).
34. Fraser, D. A. *The Physics of Semiconductor Devices* (Clarendon, Oxford, 1986).
35. Barone, A. & Paterno, G. *Physics and Applications of the Josephson Effect* (Wiley, New York, 1982).
36. Beenakker, C. W. J. *Single Charge Tunneling*, Ch. 5 (eds Grabert, H. & Devoret, M. H.) (Plenum, New York, 1992).
37. Wilkins, R., Ben-Jacob, E. & Jaklevic, R. D. *Phys. Rev. Lett.* **63**, 801 (1989).
38. Schönenberger, C. *Europhys. Lett.* (in the press).
39. Nejoh, H. *Nature* **353**, 640 (1991).
40. Pothier, H., Lafarge, P., Urbina, C., Esteve, D. & Devoret, M. H. *Physica B* **169**, 573 (1991); *Europhys. Lett.* **17**, 259 (1992).
41. Geerligs, L. J. et al. *Phys. Rev. Lett.* **64**, 2691 (1990).
42. von Klitzing, K. *Rev. mod. Phys.* **58**, 519 (1986).
43. Williams, E. R., Gosh, R. N. & Martinis, J. M. *J. Res. Natl. Inst. Stand. Technol.* **97**, (1992).
44. Averin, D. V. & Odintsov, A. A. *Phys. Lett. A* **149**, 251 (1989).
45. Averin, D. V., Odintsov, A. A. & Vyshenskii, S. V. *J. Appl. Phys.* (in the press).
46. Jensen, H. D. & Martinis, J. M. *Phys. Rev. B* **46** (in the press).
47. Pothier, H., Lafarge, P., Esteve, D., Urbina, C. & Devoret, M. H. *IEEE Trans. Magn.* (in the press).
48. Lafarge, P. et al. *C. R. Acad. Sci. Paris* **314**, 883 (1992).
49. Aviram, A. & Ratner, M. *Chem. Phys. Lett.* **29**, 277 (1974); *Molecular Electronic Devices* (ed. Carter, F. L.) (North-Holland, Amsterdam, 1991).

ACKNOWLEDGEMENTS. We thank H. Grabert, P. Joyez, P. Lafarge and H. Pothier for discussions, and P. F. Orfila for help with the figures.

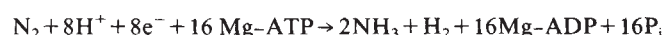
Crystallographic structure and functional implications of the nitrogenase molybdenum-iron protein from *Azotobacter vinelandii*

Jongsun Kim & D. C. Rees*

Division of Chemistry and Chemical Engineering 147-75CH, California Institute of Technology, Pasadena, California 91125, USA

The crystal structure of the nitrogenase molybdenum-iron protein from *Azotobacter vinelandii* has been determined at 2.7 Å resolution. The α - and β -subunits in this $\alpha_2\beta_2$ tetramer have similar polypeptide folds. The FeMo-cofactor is completely encompassed by the α -subunit, whereas the P-cluster pair occurs at the interface between α - and β -subunits. Structural similarities are apparent between nitrogenase and other electron transfer systems, including hydrogenases and the photosynthetic reaction centre

THE enzymatic reduction of dinitrogen to ammonia during biological nitrogen fixation is essential for maintaining the nitrogen cycle on earth. The nitrogenase enzyme system^{1–7}, which provides the biochemical machinery for nitrogen fixation, consists of two component metalloproteins, the molybdenum iron (MoFe) protein and the iron (Fe) protein. The overall stoichiometry of the biological nitrogen fixation reaction⁸:



reflects the requirements for reducing equivalents, Mg-ATP and protons, in addition to the two nitrogenase proteins. The Fe-protein is a dimer of two identical subunits coordinating a single

4Fe:4S cluster, whose crystallographic structure has been recently described⁹. The Fe-protein is initially reduced by ferredoxin or flavodoxin *in vivo*, and subsequently transfers electrons to the MoFe-protein in a process that is coupled to the hydrolysis of Mg-ATP. The MoFe-protein is an $\alpha_2\beta_2$ tetramer with a total relative molecular mass of ~240K. The two subunits are of similar size; for example the isolated α and β subunits of *Azotobacter vinelandii* MoFe-protein have 491 and 522 amino acids, respectively¹⁰. The MoFe-protein contains two types of metal centres, the FeMo-cofactor and P-cluster pair, for which structural models have been recently proposed¹¹. In this report, the tertiary and quaternary structures of the MoFe-protein from *A. vinelandii* are presented, based on a 2.7 Å resolution X-ray crystallographic analysis. The crystallographic model provides a structural basis for interpreting the functional role of the nitrogenase MoFe-protein in dinitrogen reduction.

* To whom correspondence should be addressed.

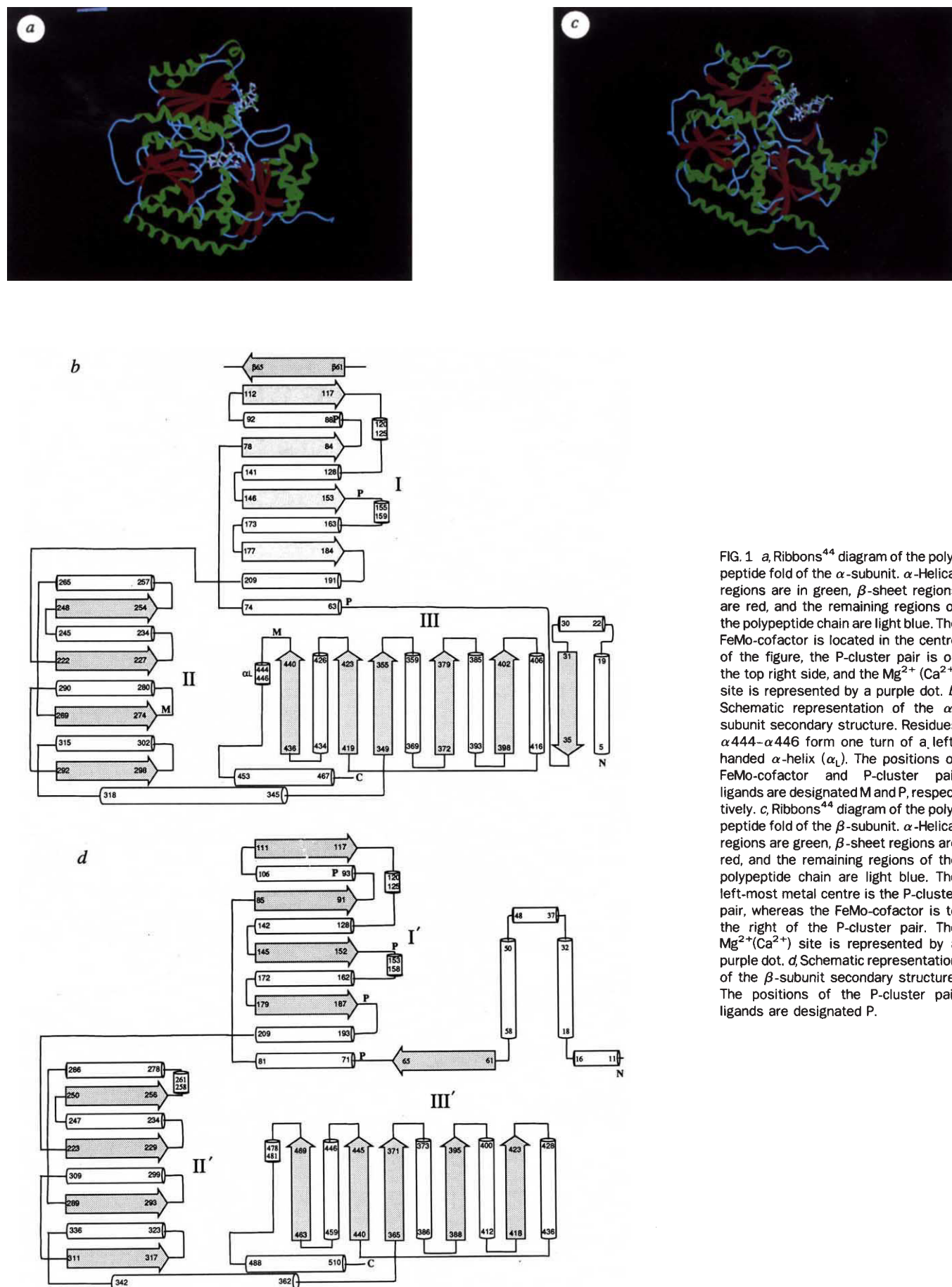


FIG. 1 *a*, Ribbons⁴⁴ diagram of the polypeptide fold of the α -subunit. α -Helical regions are in green, β -sheet regions are red, and the remaining regions of the polypeptide chain are light blue. The FeMo-cofactor is located in the centre of the figure, the P-cluster pair is on the top right side, and the Mg^{2+} (Ca^{2+}) site is represented by a purple dot. *b*, Schematic representation of the α -subunit secondary structure. Residues $\alpha 444$ – $\alpha 446$ form one turn of a left-handed α -helix (α_1). The positions of FeMo-cofactor and P-cluster pair ligands are designated M and P, respectively. *c*, Ribbons⁴⁴ diagram of the polypeptide fold of the β -subunit. α -Helical regions are green, β -sheet regions are red, and the remaining regions of the polypeptide chain are light blue. The left-most metal centre is the P-cluster pair, whereas the FeMo-cofactor is to the right of the P-cluster pair. The Mg^{2+} (Ca^{2+}) site is represented by a purple dot. *d*, Schematic representation of the β -subunit secondary structure. The positions of the P-cluster pair ligands are designated P.

Structure of the MoFe-protein

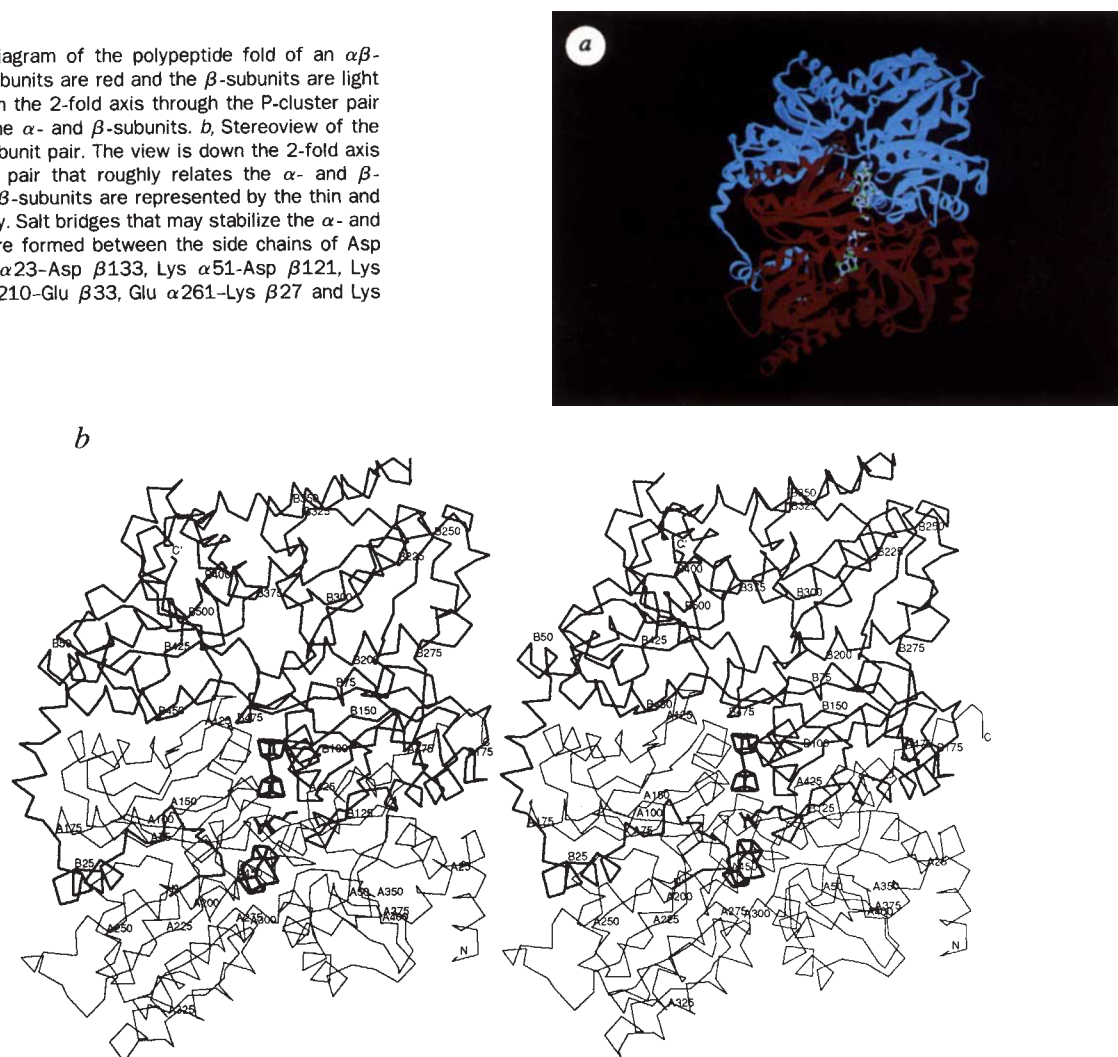
The structure determination of the nitrogenase MoFe-protein is summarized in Table 1. The α - and β -subunits exhibit similar polypeptide folds consisting of three domains (designated I, II, III and I', II', III' in the α - and β -subunits, respectively) of the α/β -type, with some extra helices (Fig. 1a-d). Similarities between the sequences of the amino-terminal regions of the α - and β -subunits had been previously noted¹². Although the α/β -type fold is a common motif in protein structure, no significant homology to other proteins has been identified for the α - and/or β -subunits by either sequence comparisons or three-dimensional/one-dimensional profile methods¹³ (D. S. Eisenberg, personal communication), with the exception of the corresponding proteins in alternate nitrogenases^{14,15} and the protein products of the *nifE* and *nifN* genes¹⁶. In each subunit, there is a wide, shallow cleft between the three domains; the FeMo-cofactor occupies the bottom of this cleft in the α -subunit. This location for the FeMo-cofactor at the boundary between three domains is reminiscent of the site for the iron-sulphur cluster in aconitase¹⁷. The site in the β -subunit corresponding to the location of the FeMo-cofactor in the α -subunit contains the side chains of residues His β 193, Gln β 294, His β 297 and Asp β 372.

Extensive contacts occur between pairs of α - and β -subunits (Fig. 2), especially in the region surrounding the P-cluster pair that bridges domain I of the α -subunit and domain I' of the β -subunit. The α - and β -subunits are roughly related by a 2-fold rotation axis that passes through the P-cluster pair. Additional

inter-subunit contacts occur between domain I of the α -subunit and domains I' and III' of the β -subunit, along with the 2-fold related set of contacts. An antiparallel β -sheet arrangement between β 61- β 65 and α 112- α 117 also contributes to the $\alpha\beta$ -subunit interface. The N terminus of the β -subunit (about 50 residues), which is absent in *Clostridium pasteurianum* MoFe protein¹⁸, extends from the β -subunit and wraps around the α -subunit.

The $\alpha_2\beta_2$ tetramer may be considered to be formed from pairs of $\alpha\beta$ dimers that are related by the molecular 2-fold rotation axis (Fig. 3a, b). Even though the α - and β -subunits in an $\alpha\beta$ dimer are also roughly related by a 2-fold rotation, the MoFe-protein does not exhibit 222 symmetry as had been earlier proposed¹⁹. At the positions of closest approach, the $\alpha\beta$ 2-fold and the tetramer 2-fold axes are separated by ~ 12 Å; furthermore the two axes are not exactly perpendicular as their orientations differ by $\sim 97^\circ$. The tetramer interface is generated by extensive interactions between domains II' and III' of the two β -subunits, along with some additional interactions involving domain III of each α -subunit. Packing between helices from the β -subunit (B234-B247, B342-B362, B323-B336, B488-B510 and the corresponding residues of subunit D) dominate interactions at the tetramer interface (Fig. 3c), with some contributions from helices in the α -subunit (A302-A315, A318-A345, A453-A467, and the corresponding residues of subunit C). These helical interactions seem to provide a major driving force for tetramerization. Intriguingly, the centre of the six α -helical barrel surrounding the tetramer 2-fold axis is not filled with side

FIG. 2 a, Ribbons⁴⁴ diagram of the polypeptide fold of an $\alpha\beta$ -subunit pair. The α -subunits are red and the β -subunits are light blue. The view is down the 2-fold axis through the P-cluster pair that roughly relates the α - and β -subunits. b, Stereoview of the α trace of an $\alpha\beta$ -subunit pair. The view is down the 2-fold axis through the P-cluster pair that roughly relates the α - and β -subunits. The α - and β -subunits are represented by the thin and thick lines, respectively. Salt bridges that may stabilize the α - and β -subunit interface are formed between the side chains of Asp α 116-Lys β 68, Lys α 23-Asp β 133, Lys α 51-Asp β 121, Lys α 76-Glu β 32, Arg α 210-Glu β 33, Glu α 261-Lys β 27 and Lys α 433-Glu β 109.



chains; rather, an open channel of $\sim 8\text{--}10\text{ \AA}$ diameter and length $\sim 35\text{ \AA}$ exists around this axis.

The tetramer interface also seems to be stabilized by cation binding to a site created by ligands from both β -subunits. The atomic identity of this site, which was initially identified as the strongest feature in difference maps, has not yet been unambiguously established, but based on the coordination environment seems to be either Mg^{2+} or Ca^{2+} . This ion has an octahedral coordination environment provided by the carboxyl oxygens of Glu B109, Asp D353 and Asp D357, the carbonyl oxygen of Arg B108 and probably two water molecules. The site is buried at the interface of the two $\alpha\beta$ dimers, and is about $25\text{ \AA}$ and $21\text{ \AA}$ away from the P-cluster pair and FeMo-cofactor, respectively. This site does not appear to have a functional role, but rather may serve to stabilize the subunit associations in the tetramer.

Environment of metal centres

The FeMo-cofactor, first identified by Shah and Brill²⁰, almost certainly provides the substrate-binding site²¹. The FeMo-cofactor and the surrounding residues are represented in Fig. 4a. The FeMo-cofactor is buried at least $10\text{ \AA}$ below the protein surface. Cys $\alpha 275$ and His $\alpha 442$, as well as Ser $\alpha 278$ which is hydrogen bonded to the S γ of Cys $\alpha 275$, are strictly conserved in all known MoFe-protein sequences. Other highly conserved residues in the vicinity of the FeMo-cofactor include: Gly $\alpha 356$ and $\alpha 357$, which are required to avoid steric interference with

the cofactor; Arg $\alpha 96$ and Arg $\alpha 359$, that can potentially form hydrogen bonds to cluster sulphurs in the FeMo-cofactor, and may serve to stabilize the FeMo-cofactor and/or partially reduced intermediates formed during substrate reduction; His $\alpha 195$, that may function in proton transfer reactions; and three residues, Gln $\alpha 191$, Glu $\alpha 440$ and Glu $\alpha 427$, which are near the homocitrate and interact with this group either directly or through water molecules. The protein environment around the FeMo-cofactor is primarily provided by hydrophilic residues, although there are some hydrophobic residues such as Tyr $\alpha 229$, Ile $\alpha 231$, Val $\alpha 70$, Phe $\alpha 381$, Leu $\alpha 358$ and Ile $\alpha 355$. The positioning of the FeMo-cofactor near the N-terminal ends of helices $\alpha 280\text{--}\alpha 290$ and $\alpha 359\text{--}\alpha 369$ may serve stabilize the cluster electrostatically.

The P-cluster pair, which may function in electron transfer between the 4Fe:4S cluster of Fe-protein and the FeMo-cofactor, is located about $10\text{ \AA}$ below the protein surface, on the 2-fold axis that roughly relates the α - and β -subunits. The P-cluster pair and surrounding residues are shown in Fig. 4b. The protein environment around the P-cluster pair is mainly provided by hydrophobic residues, including Tyr $\alpha 64$, Pro $\alpha 85$, Tyr $\alpha 91$, Pro $\alpha 155$, Phe $\alpha 186$, Pro $\beta 72$, Phe $\beta 99$, Tyr $\beta 98$, Met $\beta 154$ and Phe $\beta 189$. The location of the P-cluster pair near the N-terminal ends of six helices ($\alpha 63\text{--}\alpha 74$, $\alpha 88\text{--}\alpha 92$, $\alpha 155\text{--}\alpha 159$, $\beta 71\text{--}\beta 81$, $\beta 93\text{--}\beta 106$ and $\beta 153\text{--}\beta 158$) may serve to provide an electrostatic contribution to cluster stability²². With the exception of the cluster ligands (Cys $\alpha 62$, $\alpha 88$, $\alpha 154$, $\beta 70$, $\beta 95$, $\beta 153$

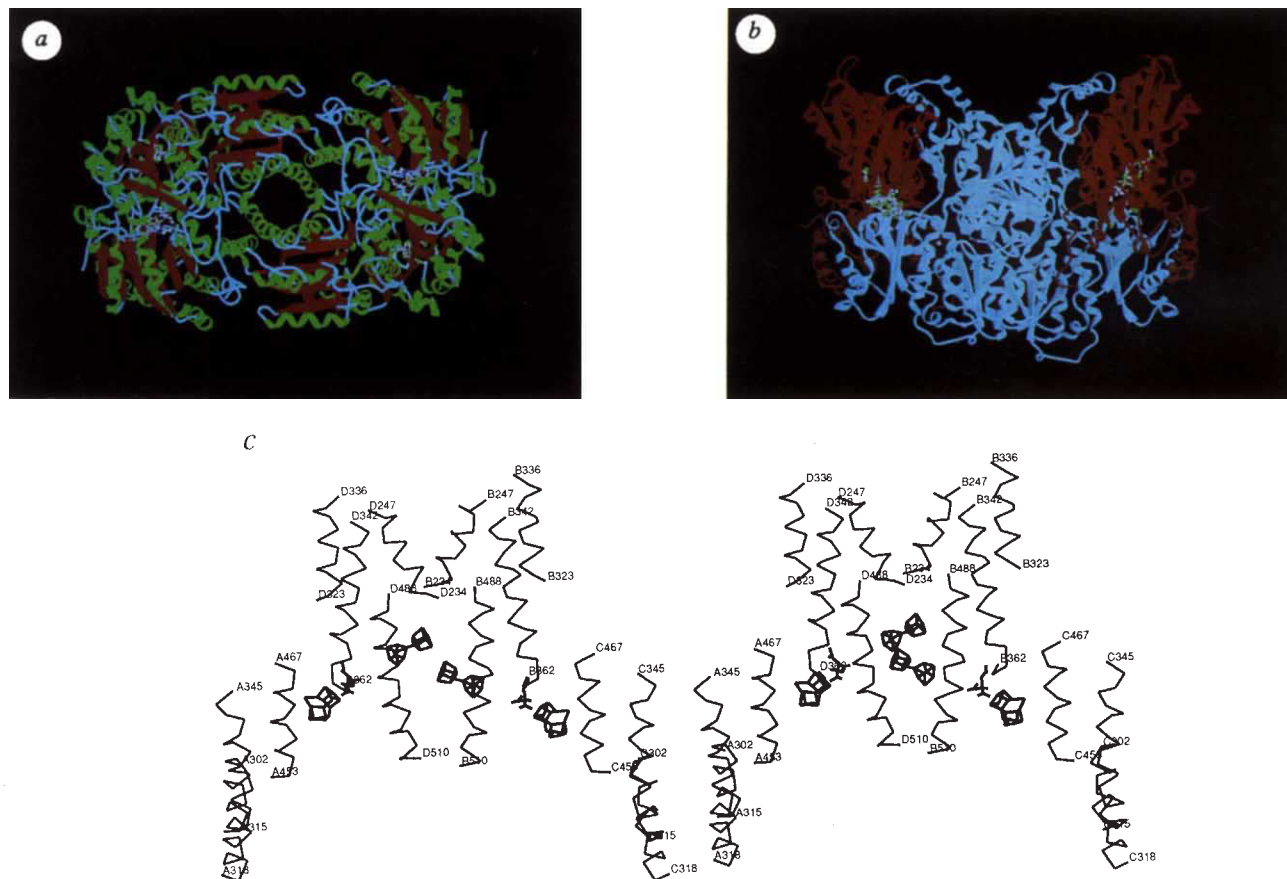
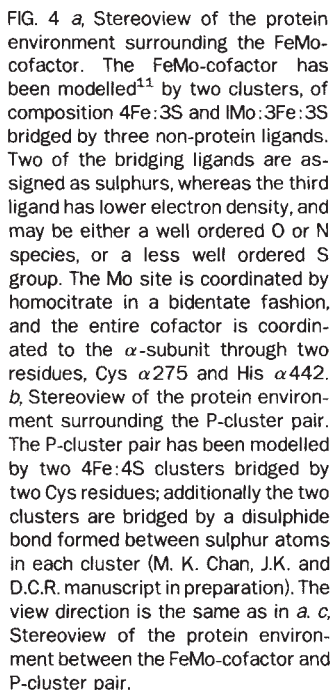


FIG. 3 a, Ribbons⁴⁴ diagram of the polypeptide fold of the $\alpha_2\beta_2$ MoFe-protein tetramer pair. The view is down the tetramer 2-fold axis. α -Helical regions are in green, β -sheet regions are red, and the remaining regions of the polypeptide chain are light blue. Salt bridges at the interface of the two $\alpha\beta$ dimers are formed between the terminal oxygen of Arg D523 and the side chains of Arg B108 and His B477; the side chains of the Arg B476-Asp D506-Lys B449 trio; Lys A433-Asp D353; Lys A474-Asp D323; and Glu

B259-Lys D346. b, Ribbons⁴⁴ diagram of the polypeptide fold of the $\alpha_2\beta_2$ MoFe-protein tetramer pair, viewed perpendicularly to the tetramer 2-fold axis. The α -subunits are red and the β -subunits are light blue. c, Stereoview of the $C\alpha$ trace of α -helices stabilizing the tetramer interface, viewed perpendicularly to the tetramer 2-fold axis. The positions of the metal centres are provided for reference.



and Ser β 188), Gln β 93 and Thr β 152, hydrophilic residues around the P-cluster pair, such as Glu α 153, Glu α 184, Ser α 92, Ser α 152 and Ser β 92 are generally not conserved in different MoFe-protein sequences. Strictly conserved Gly residues (α 87, β 94 and α 185) around the P-cluster pair are also structurally important to avoid steric interference with P-cluster pair. Substitution of Gly α 185 with Asp has been found in a Nif⁻ mutant of nitrogenase²³.

The protein environment between the FeMo-cofactor and P-cluster pair, which may be important for electron transfer between the two centres (if this actually occurs) is illustrated in Fig. 4c. The edge-edge distance of FeMo-cofactor to the P-cluster pair is about 14 Å. Four helices (α 63– α 74; α 88– α 92; α 191– α 209 and β 93– β 106) are oriented in parallel between the two metal centres and could play a role in electron transfer. In particular, the helices α 63– α 74 and α 88– α 92 adjacent to the P-cluster pair ligands Cys α 62 and Cys α 88 provide the most direct structural connection between the P-cluster pair and FeMo-cofactor. No completely covalent or hydrogen-bonded network exists along the most direct pathway between the two centres, suggesting that either some jumps must occur during electron transfer, or that structural fluctuations/alterations occur during substrate reduction that permit the formation of more favourable electron transfer pathways²⁴.

Access of substrates to active site

No permanent channels between the protein surface and the FeMo-cofactor are present in the MoFe-protein structure that could represent the entry/exit pathway for dinitrogen and ammonia. By analogy to the binding and release of oxygen to the buried haems in globins²⁵, however, it seems reasonable that structural fluctuations can create the transient cavities required for the diffusion of ligands into and away from the FeMo-cofactor, especially considering the rather slow turnover time (~ 5 s⁻¹) of nitrogenase^{26–28}. Three stretches of polypeptide chain, near residues α 192, α 277 and α 356, form a 'flap' over the FeMo-cofactor that may provide the entry/exit route for substrates and products, as well as provide a possible opening for cofactor insertion during biosynthesis of the MoFe-protein. Although covering a longer distance, it is also possible that a pathway for substrates and products could exist between the FeMo-cofactor and the protein surface lining the channel between the two β -subunits.

In addition to dinitrogen, ammonia formation also requires the delivery of protons to the active site. As no permanent channels exist by which H₃O⁺ may diffuse from the protein surface to the FeMo-cofactor, it is possible that proton transfer could use a mechanism of the type envisioned for protonation of reduced quinones in the photosynthetic reaction centre²⁹. In this mechanism, protons are transferred by shuttling from one functional group to another; the side chains of Asp, Glu and His residues, as well as water molecules, would seem to be particularly suited for this function. Two patches of histidines occur on the α -subunit surface that might represent the initial site for proton binding; these patches contain the imidazole side chains of α 196 and α 383; and α 274, α 362 and α 451. But because several of these residues have been substituted either in naturally occurring MoFe-proteins or by site-directed mutagenesis without serious loss of activity³⁰, either they are not involved in proton transfer, or else there is a multiplicity of proton transfer pathways that can tolerate some alterations. In the interior of the protein, several potential pathways occur that could funnel protons to FeMo-cofactor: Glu α 440 to homocitrate to FeMo-cofactor, and His α 195 to FeMo-cofactor. If an intermediate that is a sufficiently strong base is generated during dinitrogen reduction, it is possible that protons may also be transferred along Asp α 234 to Arg α 359 to the FeMo-cofactor, or by Arg α 96 to FeMo-cofactor. The presence of multiple, potential proton transfer routes suggests that there is

TABLE 1 Structure determination and refinement statistics

(a) Heavy atom binding sites in crystals of <i>A. vinelandii</i> MoFe-protein					
Derivative	x	y	z	Relative occupancy	Location
EMTS	0.144	-0.265	0.124	0.41	Cys α 472
	0.423	0.000	0.124	0.39	Cys α 472
					NCS related
	-0.198	-0.272	0.029	0.14	Cys α 45
	0.714	0.008	0.028	0.14	Cys α 45
					NCS related
	-0.172	-0.335	-0.352	0.11	Cys α 324, Glu α 325
	0.487	0.070	-0.347	0.06	Cys α 324, Glu α 325
					NCS related
PtCl ₄	0.292	-0.321	0.323	0.36	Met β 330
	0.377	0.061	0.314	0.39	Met β 330
					NCS related
	0.250	-0.167	0.184	0.33	Met β 491
	0.350	-0.092	0.190	0.30	Met β 491
					NCS related
	-0.198	-0.277	0.022	0.28	Met α 391
	0.711	0.018	0.016	0.25	Met α 391
					NCS related
	-0.240	-0.271	-0.293	0.11	Met α 295, Lys α 315
	0.583	0.008	-0.300	0.13	Met α 295, Lys α 315
					NCS related
	-0.024	-0.335	0.209	0.11	Met α 4?
					(not identified yet)
	0.618	0.067	0.171	0.11	Met α 4?
					(not identified yet)
					NCS Related

PIP has seven common binding sites with the PtCl₄ derivative

(b) Refinement statistics

Model status		
Number of amino-acid residues	1980/2026	
Total non-hydrogen atoms	15,858 (98% complete)	
Number of water molecules	4 (ligands to divalent cation sites)	
Missing residues	α 2– α 4, α 36– α 44, α 482– α 492	
Number of Ramachandran outliers	10/1980	
Refinement statistics		
	Before refinement	After refinement (X-PLOR)
R-factor (8.0–2.7 Å)	0.368	0.188
R.M.S. deviation of bond lengths (Å)	0.020	0.016
bond angles (°)	2.77	3.29
dihedral angles(°)	27.95	24.15
improper torsion(°)	1.50	1.41

As described in ref. 11, three crystal forms were prepared of the MoFe-protein from *A. vinelandii* (abbreviated Av1) and *Clostridium pasteurianum*, each containing one tetramer per asymmetric unit. The most detailed structural analysis has been done for an Av1 crystal form, space group P2₁ with unit cell constants of $a=108.4$ Å, $b=130.5$ Å, $c=81.5$ Å, and $\beta=110.8^\circ$. Heavy-atom derivatives were prepared¹¹ with ethylmercurithiosalicylate (EMTS), K₂PtCl₄ (PtCl₄) and di- μ -iodobis(ethylenediamine)-diplatinum(II) nitrate (PIP). Locations for the heavy atoms were determined from difference Patterson maps (Table 1a), and were used to calculate multiple isomorphous replacement (MIR) phases with the program HEAVY³⁸. Phases were refined by non-crystallographic symmetry (NCS) averaging both within and between the different crystal forms^{39,40}. The protein model was built with the graphics program TOM/FRODO⁴¹. The electron density map was improved by iterative cycles of model building, refinement, phase combination and NCS averaging. During successive cycles of phase refinement, the resolution was gradually extended from 3.5 Å to 2.7 Å. The Av1 native data set contains 53,736 reflections, and is 94% complete to 2.75 Å resolution¹¹. The program TNT⁴² was used for coordinate refinement during the initial stages of modelling, whereas the simulated annealing program X-PLOR⁴³ was used during the final stages of refinement with the PARAM19X.PRO parameter file. Residues were rebuilt following inspection of 2F_o–F_c maps, and examination of Eisenberg's three-dimensional-one-dimensional profile^{1,3} calculated for each model. Final refinement statistics are presented in Table 1b. Residue numbers of the MoFe-protein are prefixed with either α or β to indicate the appropriate subunit number, unless interactions in the tetramer are described, in which case the prefixes A and C designate the two distinct α subunits, while the prefixes B and D designate the two distinct β subunits present in the tetramer. No subunit prefix is included for residues in the nitrogenase Fe-protein.

not a unique pathway by which protons are shuttled from the surface to the active site.

Interaction between MoFe-protein and Fe-protein

Two residues of Fe-protein that have been identified as interacting with the MoFe-protein, Arg 100³¹ and Glu 112³², are located

on the same side of the Fe-protein as the 4Fe:4S cluster⁹. Hence this surface almost certainly includes at least part of the interaction region between the two proteins. A plausible model¹¹ for docking the two proteins involves superposition of the Fe-protein 2-fold axis with the pseudo 2-fold axis passing through the P-cluster pair that relates the α - and β -subunits of the MoFe-protein (Fig. 5a). To either side of the P-cluster pair are two wide and shallow clefts related by the pseudo 2-fold axis (Fig. 2), that could accommodate the two Fe-protein subunits. With this orientation, the side chains of Glu 112 and Lys β 400 can be positioned sufficiently close to permit the experimentally observed crosslinking³².

The potential 4Fe:4S cluster binding site and the surrounding residues are shown in Fig. 5b. Four short helices (α 155– α 159; α 120– α 125; β 153– β 158 and β 120– β 125) that are related by a pseudo 2-fold axis, are oriented in parallel from the P-cluster pair toward the surface forming a four-helical bundle, and the 4Fe:4S cluster of Fe-protein could bind to the top surface of these helices. Kinetic studies of site-directed mutants at Phe β 125 indicate that these alterations interfere with the Fe-protein/MoFe-protein interaction³³. Additionally, substitutions of Glu α 120 and Gly α 160 in or near these helices have been identified in Nif[−] mutants²³, possibly due to disruption of the Fe-protein/MoFe-protein interaction. The edge–edge distance from the P-cluster pair to the 4Fe:4S cluster of Fe-protein in this model would be about 15 Å. The hydrophobic residues Phe α 125, Phe β 125, Phe α 186, Phe β 189, Leu α 158, Val β 157, Ile β 158 and Ile β 123 contribute to the interface between the P-cluster pair and the proposed Fe-protein cluster binding site, and they are generally conserved in different MoFe-protein sequences. Charged and polar residues on the MoFe-protein surface in this region may be important for the formation of

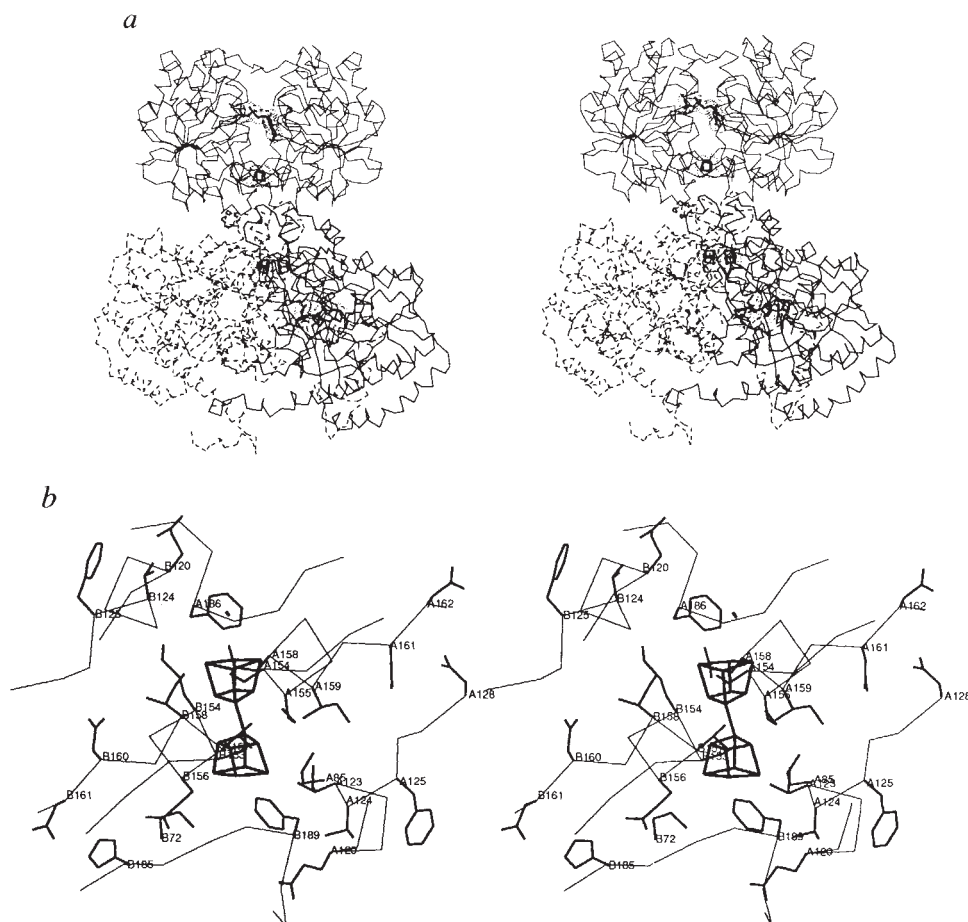
salt bridges/hydrogen bonds with Fe-protein. Asp α 162 and Asp β 161, which are related by a pseudo-2-fold axis, are completely exposed to the solvent and are located about 15 Å away from the P-cluster pair. Other solvent-exposed side chains near the Fe-protein binding site include His β 185, Asp α 128, Asp α 161 and Glu β 156.

Kinetic studies indicate that nucleotide hydrolysis precedes electron transfer in the Fe-protein/MoFe-protein complex^{26–28}. The nucleotide-binding sites in the Fe-protein dimer are located at the interface between the two subunits⁹. Although the detailed structural consequences of nucleotide binding on the Fe-protein structure have not been established, it is likely that Mg-ATP hydrolysis is accompanied by a change in Fe-protein structure, such as an alteration in the relative orientations of the two subunits. Presumably, this transition leads to the formation of an activated species that is competent for electron transfer from the Fe-protein to MoFe-protein (in particular, to the P-cluster pair in this model). As the nucleotide-binding site is on the opposite surface of the Fe-protein to the interaction regions with the MoFe-protein, it should be possible for Mg-ATP to exchange with Mg-ADP without dissociation of the two proteins, as has been proposed from kinetic analyses of the nitrogenase reaction²⁸.

Similarities with other electron transfer systems

Hydrogenases. Hydrogen evolution is an obligatory part of the nitrogenase mechanism; in addition, in the absence of other reducible substrates, the total electron flux through nitrogenase is funnelled into hydrogen production³⁴. Consequently, nitrogenase can be considered as a hydrogenase, and there may be similarities in the structures of the catalytic metal centres found in these two classes of enzymes. The active site of iron-only

FIG. 5 a, Stereoview of C α model for a complex formed between Fe-protein (thin lines) and an $\alpha\beta$ subunit pair from the MoFe-protein (dark lines, with the polypeptide fold of the β -subunit represented by dashed lines). The metal centres in the two proteins are represented by models surrounded by a dotted van der Waals surface. This model was generated by graphical superimposition of the crystal structures of the individual proteins. b, Stereoview of the protein environment in the putative Fe-protein binding site, near the MoFe-protein P-cluster pair.



hydrogenases is provided by a metal centre (the 'H-cluster') that contains ~6 Fe and ~6S, with some non-sulphur ligation³⁵. It is possible that the construction of the FeMo-cofactor from bridged clusters, as seen in the nitrogenase MoFe-protein, may be relevant to the molecular structure of the H-cluster³⁵.

Photosynthetic reaction centre. Although the electron transfer processes catalysed by nitrogenase and the photosynthetic reaction centre (RC; reviewed in ref. 29) are quite different, there are striking similarities in the structural organization of the two systems. (1) Both the MoFe-protein and the RC are composed of two homologous subunits roughly related by a 2-fold rotation: the α and β subunits of the MoFe-protein¹¹, and the L and M subunits of the reaction centre^{36,37}. The initial electron carriers for both systems, the special pair in the RC and (most likely) the P-cluster pair in the MoFe-protein, are buried at the interface between the two subunits. The location of these redox centres at the subunit interface may provide a convenient assembly mechanism for the incorporation of powerful reductants in the protein interior, isolated from contact with solvent. (2) Despite the general 2-fold symmetry in the protein organization, electron transfer from the initial donor proceeds in only one direction. In the RC, only one of the two electron transfer branches present is actually used. In the MoFe-protein, however, the FeMo-

cofactor is present in only one of the two homologous subunits, so that electron transfer necessarily proceeds in this direction. Whether the original MoFe protein was a homodimer or tetramer containing equivalent branches is an interesting question in molecular evolution. (3) The binding sites for the terminal electron acceptors (dinitrogen and quinone) are buried in both the MoFe-protein and RC, so that the protein structures must accommodate entry and exit of substrates and products. (4) In the RC, it has been established that separate pathways exist for electron transfer and proton transfer within the protein, and it seems likely that the MoFe-protein behaves similarly. (5) a significant fraction of the energy input into these systems (light for the RC and Mg-ATP for nitrogenase) is lost during the overall reactions. This may be as a consequence of ensuring that the electron transfer reactions are essentially irreversible, so that short-circuits or futile cycles are minimized.

Despite the accomplishments over the past century in characterizing the biochemical, spectroscopic and genetic properties of nitrogenase, a detailed molecular description of the enzyme mechanism has been elusive, in part due to the absence of structural information. The crystallographic structures of the nitrogenase proteins should provide the necessary framework for addressing outstanding issues related to the assembly and mechanism of the nitrogenase system. □

Received 17 September; accepted 24 November 1992.

- Burgess, B. K. in *Advances in Nitrogen Fixation* (eds Veege, C. & Newton, W. E.) 103-114 (Martinus Nijhoff, Boston, 1984).
- Orme-Johnson, W. H. A. *Rev. Biophys. biophys. Chem.* **14**, 419-459 (1985).
- Holm, R. H. & Simhon, E. D. in *Molybdenum Enzymes* (ed. Spiro, T. G.) Chapter 1 (Wiley-Interscience, New York, 1985).
- Stiefel, E. I. et al. in *Metal Clusters in Proteins* (ed. Que, L.) 372-389 (American Chemical Society, Washington DC 1988).
- Burgess, B. K. *Chem. Rev.* **90**, 1377-1406 (1990).
- Burris, R. H. *J. biol. Chem.* **266**, 9339-9342 (1991).
- Smith, B. E. & Eady, R. R. *Eur. J. Biochem.* **205**, 1-15 (1992).
- Simpson, F. B. & Burris, R. H. *Science* **224**, 1095-1097 (1984).
- Georgiadis, M. M. et al. *Science* **257**, 1653-1659 (1992).
- Brigle, K. E., Newton, W. E. & Dean, D. R. *Gene* **37**, 37-44 (1985).
- Kim, J. & Rees, D. C. *Science* **257**, 1677-1682 (1992).
- Lammers, P. J. & Haselkorn, R. *Proc. natn. Acad. Sci. U.S.A.* **80**, 4723-4727 (1983).
- Bowie, J. U., Lüthy, R. & Eisenberg, D. *Science* **253**, 164-170 (1991).
- Bishop, P. E. et al. in *Nitrogen Fixation: Hundred Years After* (eds Bothe, H., deBruijn, R. J. & Newton, W. E.) 71-79 (Gustav Fischer, Stuttgart, 1988).
- Eady, R. R. *Adv. inorg. Chem.* **36**, 77-102 (1991).
- Brigle, K. E., Weiss, M. C., Newton, W. E. & Dean, D. R. *J. Bact.* **169**, 1547-1553 (1987).
- Robbins, A. H. & Stout, C. D. *Proteins Struct. Funct. Genet.* **5**, 289-312 (1989).
- Wang, S.-Z., Chen, J.-S. & Johnson, J. L. *Biochemistry* **27**, 2800-2810 (1988).
- Yamane, T., Weininger, M. S., Mortenson, L. E. & Rossmann, M. G. *J. biol. Chem.* **257**, 1221-1223 (1982).
- Shah, V. K. & Brill, W. *proc. natn. Acad. Sci. U.S.A.* **74**, 3249-3253 (1977).
- Hawkes, T. R., McLean, P. A. & Smith, B. E. *Biochem. J.* **217**, 317-321 (1984).
- Adman, E. T., Watenpaugh, K. D. & Jensen, L. H. *Proc. natn. Acad. Sci. U.S.A.* **72**, 4854-4858 (1975).
- Govenzensky, D. & Zamil, A. *J. Bact.* **171**, 5729-5735 (1989).
- Wuttke, D. S., Bjerrum, M. J., Winkler, J. R. & Gray, H. B. *Science* **256**, 1007-1009 (1992).
- Case, D. A. & Karplus, M. *J. molec. Biol.* **132**, 343-368 (1979).

- Lowe, D. J. & Thorneley, R. N. F. *Biochem. J.* **215**, 393-405 (1983).
- Lowe, D. J. & Thorneley, R. N. F. *Biochem. J.* **224**, 895-901 (1984).
- Thorneley, R. N. F. in *Nitrogen Fixation: Achievements and Objectives* (eds Gresshoff, P. M., Roth, L. E., Stacey, G. & Newton, W. E.) 103-109 (Chapman and Hall, New York, 1990).
- Feher, G., Allen, J. P., Okamura, M. Y. & Rees, D. C. *Nature* **339**, 111-116 (1989).
- Thomann, H., Bernardo, M., Newton, W. E. & Dean, D. R. *Proc. natn. Acad. Sci. U.S.A.* **88**, 6620-6623 (1991).
- Murrell, S. A., Lowery, R. G. & Ludden, P. W. *Biochem. J.* **251**, 609-612 (1988).
- Willing, A. & Howard, J. B. *J. biol. Chem.* **265**, 6596-6599 (1990).
- Thorneley, R. N. F., Ashby, G. A., Fisher, K. & Lowe, D. J. in *Molybdenum Enzymes, Cofactors and Models* (eds Stiefel, E., Coucouvanis, D. & Newton, W. E.) (American Chemical Society, Washington DC in the press).
- Hadfield, K. L. & Bule, W. A. *Biochemistry* **8**, 5103-5108 (1969).
- Adams, M. W. W. *Biochim. biophys. Acta* **1020**, 115-145 (1990).
- Deisenhofer, J., Epp, O., Miki, K., Huber, R. & Michel, H. *Nature* **318**, 618-624 (1985).
- Allen, J. P., Feher, G., Yeates, T. O., Komiya, H. & Rees, D. C. *Proc. natn. Acad. Sci. U.S.A.* **84**, 5730-5734 (1987).
- Terwilliger, T. C., Kim, S.-H. & Eisenberg, D. *Acta crystallogr.* **A43**, 1-5 (1987).
- Rossmann, M. G. *Molecular Replacement Method* (Gordon and Breach, New York, 1972).
- Bricogne, G. *Acta crystallogr.* **A32**, 832-847 (1976).
- Jones, T. A. *Meth. Enzym.* **115**, 151-171 (1985).
- Tronrud, D. E., Ten Eyck, L. F. & Matthews, B. W. *Acta crystallogr.* **A43**, 489-501 (1987).
- Brünger, A. T. *J. molec. Biol.* **203**, 803-816 (1988).
- Carson, M. & Bugg, C. E. *J. molec. Graphics* **4**, 121-122 (1986).

ACKNOWLEDGEMENTS. We thank J. B. Howard, A. J. Chirino, B. T. Hsu, M. K. Chan, H. Komiya, D. Woo, G. Santillan and D. Malerba for contributions to this project. This work was supported by the NSF, with instrumentation funded in part by the Beckman Institute and the Joseph Irvine Equipment fund. X-PLOR calculations were done on the CRAY-YMP at the San Diego Supercomputer Center, supported by the NSF. Coordinates will be deposited in the Brookhaven Protein Data Bank, and are available by E-mail (REES@CITRUS.CALTECH.EDU).

The O/C abundance ratio in absorbing gas clouds at high redshift

D. Reimers*, S. Vogel*, H.-J. Hagen*, D. Engels*,
D. Groote*, W. Wamsteker†‡, J. Clavel‡§
& M. R. Rosa‡||

* Hamburger Sternwarte, Universität Hamburg, Gojenbergsweg 112,
2050 Hamburg 80, Germany

† ESA IUE Observatory, Apartado 54065, Madrid, Spain

§ ESA/ESTEC, P.O. Box 299, NL-2200 AG Noordwijk, The Netherlands

|| ST-ECF/ESO, Karl-Schwarzschild-Str. 2, 8046 Garching, F.R. Germany

THE intrinsic extreme ultraviolet (EUV) spectrum of quasars, in which spectral lines of He and CNO ions should appear, is in most cases strongly absorbed by the hydrogen Lyman continuum of intergalactic matter at cosmological distances. The bright quasar HS1700+6416 (at redshift $z=2.72$), from the Hamburg survey¹, is a fortuitous exception, the line of sight towards it being relatively transparent. Here we report detailed ultraviolet spectroscopic measurements of HS1700+6416, obtained with the Hubble Space Telescope. We observe a rich absorption line spectrum, in which many EUV resonance lines (including He I, O II to O V, and N III and N IV) are seen. The O/C abundance ratios inferred from the line-strengths are higher than Solar System values by factors of 3 to 5. This suggests that the absorbing material is halo gas of early galaxies, from which population II stars have yet to form, oxygen being a primary nucleosynthesis product from a first generation of massive stars.

The bright ($V=16.1$) quasar HS1700+6416 is an ideal candidate for a study of absorption line systems with the Hubble Space Telescope (HST) because of its remarkably unattenuated continuum as seen by the International Ultraviolet Explorer (IUE) at low signal-to-noise ratio¹. Spectra were obtained in the range 120 to 330 nm with the Faint Object Spectrograph (FOS)³ on the HST at a resolution of $\lambda/\Delta\lambda=1,300$ and a typical signal to noise ratio of ~ 20 for wavelength $\lambda > 170$ nm in a total of 11 hours exposure time. Initial inspection of Fig. 1, a composite of the three FOS wavelength ranges, shows that the spectrum is characterized by a sequence in redshift of seven optically thin Lyman Limit Systems (LLSs) which can easily be identified with strong Ly α lines and corresponding higher Lyman series members (up to L_{12}); the spectrum looks like a textbook illustration for continuous opacity. The observed run of hydrogen continuum absorption with wavelength shows a minimum of transparency at $\sim 2,500$ Å, the 'Lyman valley', which was predicted by statistical models of the distribution of matter at cosmological distance along the line of sight². The same models also predict that, for example at redshift 2.85, the probability of finding a transparent line of sight like that for HS1700+6416 is only $\sim 15\%$ (P. Møller, personal communication).

Superimposed on the continua between the LLS edges are several hundred absorption lines. Most of these can be identified with EUV lines from abundant elements in 10 different absorption line systems, seven of which are LLSs. There are also a few galactic interstellar lines at zero redshift. As an example we show the range 2,050 to 2,300 Å with prominent lines identified (Fig. 2). We identify the broad quasar emission line centred around 2,115 Å with the Ne V line at 568.4 Å. Other possible quasar emission lines are indicated in Fig. 1. The He I resonance line 584.3 Å is not seen in emission. The interpretation of the Lyman absorption edges in terms of hydrogen column densities is straightforward, as all LLSs are optically thin. Together with

TABLE 1 Lyman limit and metal line systems

Redshift z	$\log N_{\text{H}}$	$N_{\text{H I}}/N_{\text{He I}}$	Observed ions	[O/C]	$\tau(10^9 \text{ yr})$
2.433	16.7		He I, O V, IV, II, N IV, III, C IV, III, Si IV, III	0.45	13.9
2.315	16.85	60	He I, O V, IV, III, N V, IV, C IV, III, II, Si IV	0.55	13.7
2.1678	16.85	30	He I, O V, IV, C IV, III, Al II	0.7	13.4
1.8465	16.75	500	He I, O V, IV, III, O IV, III	0.5	12.7
1.725	17.05		O IV, III, O III, N III		
1.1572	16.85		O V, IV, III, O III, S III	-0.25	10.5
0.8642	16.35		C III, Si III, S III, N III, O VI		
2.579	no LL (15.6†)		He I, O V, IV, C IV, III	0.7	14.1
2.440	no LL (16.1*)	25	He I, O V, IV, C III	0.75	13.9
2.308	no LL (15.7*)		He I, O V, N V, IV, C IV, III, S VI	0.5	13.7

[O/C] = $\log N_{\text{O}}/N_{\text{C}} - 0.25$ (ref. 9). τ , look back time for $H_0=50$, $q_0=0$.
Friedman time is $\tau_F=19.6 \times 10^9$ years.

* Column densities based on Ly α and Ly β .

† Based on Ly α only.

the known frequency (ν) dependence of the Lyman absorption coefficient ($\sim \nu^{-3}$) the intrinsic quasar flux distribution and the hydrogen column densities N_{H} can be determined accurately and uniquely (Table). The redshift distribution of the seven LLSs is consistent with that found previously from the distribution of many single optically thick LLSs per line of sight⁴⁻⁶.

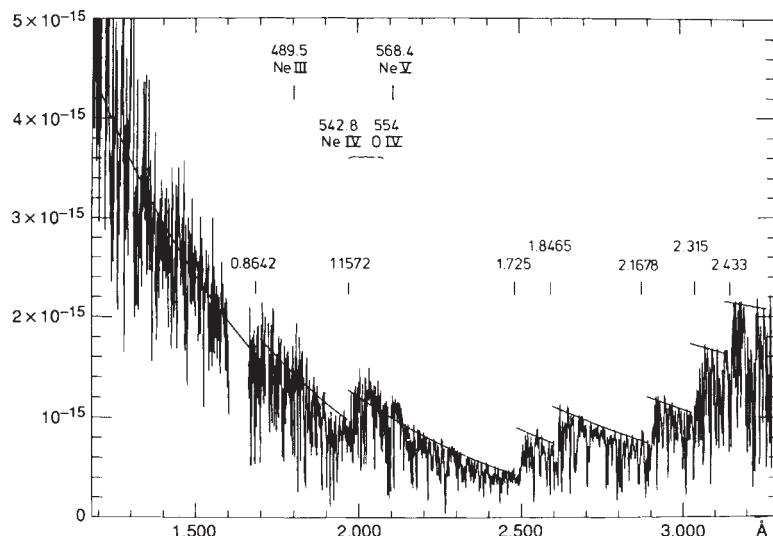
The reconstructed intrinsic flux of HS1700+6416 is characterized by a rather flat distribution between 3,270 and $\sim 2,000$ Å with an upturn to shorter wavelengths. The unique observation in HS1700+6416 is, however, the absorption line spectrum. All seven LLSs with precise redshifts from the corresponding Lyman series contain metals. In addition, we find three strong Ly α , Ly β systems that contain metals with similar absorption line spectra as the LLSs. Many lines remain unidentified either because the available EUV atomic data are incomplete, or because of Ly α forest lines. Prominent new quasar absorption lines, which have not been seen previously in any cosmic object except the Sun, are O V 629.7 Å, O IV 787.7, 608.4, 554.1, 553.3, N IV 765.1 and C III 386.3. C III at 977 Å is always prominent. The He I lines 584.33 and 537.03 Å are seen in all absorption line systems covered by our spectra (Table 1). There is no convincing evidence for He I in pure Ly α forest clouds.

Column densities for He I and metal ions are derived using the 'curve of growth' technique whereby the strength of an absorption line is tied to an empirical relation between line strengths and column densities observed for the Lyman series members of LLSs. Neutral H/He ratios thus derived range between 25 and 500, compared with a total H/He ratio of 13 in primordial matter. The excess ionization of He over H is probably due to a hard, nonstellar background radiation field of quasars^{7,8}. The ionization of He in LLS clouds contains information on the ionizing background radiation field in the early Universe. Here, we will just note that the He II Ly α 304 Å forest is probably stronger than the H Ly α forest, and that radiation from HS1700+6416 might be heavily absorbed by He II below 1,130 Å. This could be tested in future space experiments.

We next turn to the rich metal absorption-line spectrum. Previous studies of the metal content of LLS^{7,8} used ions such as C IV and Si IV, which require large ionization corrections. As the ionizing radiation fields at high redshift are poorly known, metal abundances in intergalactic clouds are uncertain. Even the nature of the objects responsible for the absorption line

‡ Affiliated to the Astrophysics Division, Space Science Department, European Space Agency.

FIG. 1 Spectrum of HS1700+6416 taken with the Faint Object Spectrograph of the Hubble Space Telescope. The resolution is $\lambda/\Delta\lambda=1,300$. Note the seven optically thin Lyman continuum jumps, the 'Lyman limit' systems with redshifts between $z=0.864$ and $z=2.433$. Superimposed on the 'steps' are several hundred quasar absorption lines. New quasar emission lines are also indicated. The wavelength dependence of the transparency of intergalactic absorption shows the 'Lyman valley' as predicted². The plotted step structure (solid lines) is calculated from a nearly flat quasar continuum ($f_{\lambda} \sim \lambda^{-0.7}$) for $\lambda > 2,000 \text{ \AA}$ with an upturn for $\lambda < 2,000 \text{ \AA}$ ($f_{\lambda} \sim \lambda^{-1.25}$), combined with hydrogen continuum absorption with column densities given in Table 1.

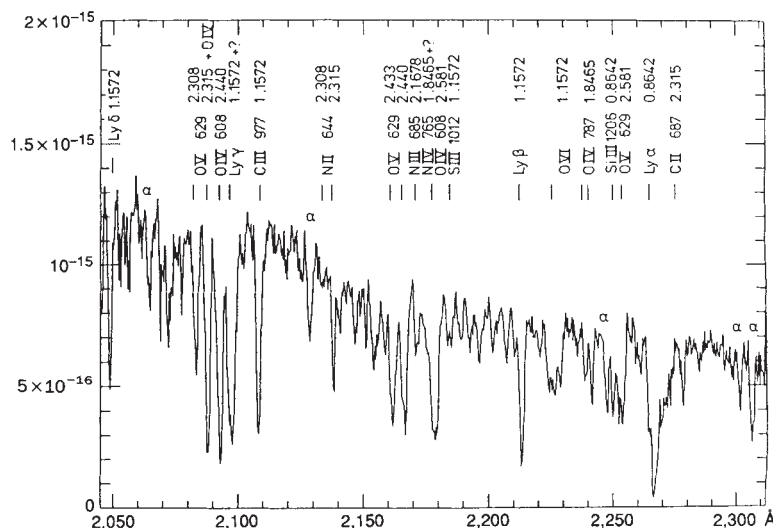


systems is not fully established. Because in the spectrum of HS1700+6416 all relevant ionization stages of oxygen (III, IV, V and sometimes VI), carbon (III, IV, II) and nitrogen (IV, III, V) are directly observable, the abundance ratios O/C and N/C can be determined without ionization calculations for most redshift systems and without knowing the absolute metal abundances. Errors in these estimates, possibly introduced by blends with unidentified absorption lines, are largely eliminated either by comparison between results from several lines of the same ion or, in cases where a particular ion is observed in only one line (O V 629.7 Å, N IV 765.14 Å), by comparison with neighbouring ionization stages. Unresolved blends in the C III line at 977 Å (the dominant C ion) mimic low relative oxygen abundances, contrary to what is actually observed. A final test on unresolved blends was made by comparing ionization calculations with observed column density ratios. Models that fit the H I/He I column density ratios also predict correctly the ratios O III/O IV/O V, C III/C IV and N III/N IV or N IV/N V. As a result we further obtain a good first estimate of the metal abundance $[C/H] = \log N_C/N_H - \log (N_C/N_H)_{\text{solar}}$. We find $-1.5 \geq [C/H] \geq -2.7$ in all systems. Carbon has been used here as an indicator of 'metal abundance' instead of Fe, which is not observable, because $[C/Fe]=0$ for halo stars with metals deficient to at least $[Fe/H]=-2$ (ref. 9). The most important

new result is that the O/C ratios are enhanced by factors between 3 and 5 (Table 1) in all systems of high redshift ($z \geq 1.8465$). In one system ($z = 2.308$) we find N/C also enhanced by a factor of 12, and in two others (2.315, 1.8465) there is also evidence for N enhancement by factors of 10 and 4, respectively, whereas in most cases N/C is at the cosmic value or below.

The O/C and N/C enrichment is similar to values found in unevolved extremely metal-deficient population II stars⁹: O is enriched by factors up to 5 relative to C and Fe in all stars, and N has been found enriched by large factors in a few (4) stars¹⁰. This is further evidence that LLSs and metal line systems at high redshift are extremely metal-deficient clouds in the halos of young galaxies containing the products of nucleosynthesis of early stars. Population II stars are later formed from these. For low redshift ($z < 1.3$) LLSs, galaxy cross-sections of the order of 100 kpc have been estimated¹¹. While the high primary production of oxygen probably originates in an initial mass function enhanced in massive stars in the first stellar generations¹²⁻¹⁴, the origin of N is less clear. Because N and O abundances are not closely correlated, the primary source for N must be different. Among other possibilities, very massive (500 solar masses) zero-metallicity stars have been proposed¹⁵. In any case, halo clouds seem to be chemically inhomogeneous at early epochs. A further implication of cosmological relevance

FIG. 2 Small section of the spectrum in Fig. 1, with identifications. The signal to noise ratio is about 20. Note the broad Ne V 568.4 Å line centred at 2,115 Å. The label α designates as yet unidentified lines with rest equivalent width $W_{\lambda, \text{rest}} \geq 0.3 \text{ \AA}$, which may be Ly α forest lines.



is that an enhanced oxygen abundance in extremely metal-deficient globular clusters such as M15 would lead to lower ages as inferred from the fitting of isochrones, namely 14 (16) Gyr for $[O/Fe] = 1(0.5)$ instead of 18 Gyr for $[O/Fe] = 0$ (ref. 16).

The good signal-to-noise ratio of the FOS spectra near 2,174 Å, the rest wavelength of the He I line at 584.33 Å, provides a more sensitive He I Gunn-Peterson test¹⁷ of the intergalactic medium than previous tests¹⁸. We find $n_{\text{HeI}} \leq 7 \times 10^{-12} \text{ cm}^{-3}$ at $z = 2.72$ ($q_0 = 0$, $H_0 = 50$), a factor of 7 lower than previous limits¹⁸.

The influence of a decaying massive neutrino¹⁹ on the softness of the background ionizing radiation field and its consequences for the visibility of a He I 584 Å forest and of He I 504 Å continuum absorption in the ultraviolet spectrum of HS1700+6416 has been discussed elsewhere²⁰. The present observations strengthen the conclusion²⁰ that the lack of visibility of either phenomenon in HS1700+6416 rules out the existence of a massive decaying neutrino as a chief component of dark matter. □

Received 13 July; accepted 20 October 1992.

1. Reimers, D. *et al.* *Astr. Astrophys.* **218**, 71–77 (1989).
2. Möller, P. & Jakobsen, P. *Astr. Astrophys.* **228**, 299–309 (1990).
3. Ford, H. C. *Faint Object Spectrograph Handbook* (Baltimore, Space Telescope Science Institute, 1985).
4. Tytler, D. *Nature* **298**, 427–432 (1982).
5. Sargent, W. L. & Steidel, C. C. *Astrophys. J. Suppl.* **69**, 703–761 (1989).
6. Lanzetta, K. M. *Astrophys. J.* **375**, 1–14 (1991).
7. Chaffee, F. H., Foltz, C. B., Bechtold, J. & Weymann, R. O. *Astrophys. J.* **301**, 116–123 (1986).
8. Steidel, C. C. *Astrophys. J. Suppl.* **74**, 37–91 (1990).
9. Gehru, T. in *Rev. mod. Astr.* **1**, 52–102 (1988).
10. Barbuy, B. *Astr. Astrophys.* **123**, 1–6 (1983).
11. Bahcall, J. N. in *Science with the Hubble Space Telescope* (ESA-NASA Workshop, in the press).
12. Arnett, W. D. *Astrophys. J.* **219**, 1008–1016 (1978).

13. Maeder, A. *Astr. Astrophys.* **92**, 101–110 (1981).
14. Clegg, R. E. S., Lambert, D. L. & Tomkin, J. *Astrophys. J.* **250**, 262–275 (1981).
15. Laird, J. B. *Astrophys. J.* **289**, 556–569 (1985).
16. VandenBerg, D. A. in *Production and Distribution of C, N, O Elements* (eds Danziger, I. J., Matteucci, F. & Kjær, K.) 73–81 (ESO, 1985).
17. Gunn, J. E. & Peterson, B. A. *Astrophys. J.* **142**, 1633–1636 (1965).
18. Tripp, T. M., Green, R. F. & Bechtold, J. *Astrophys. J.* **364**, L29–L32 (1990).
19. Sciama, D. W. *Astrophys. J.* **364**, 549–554 (1990).
20. Miralda-Escudé, J. & Ostriker, J. P. *Astrophys. J.* **392**, 15–22 (1992).

ACKNOWLEDGEMENTS. We thank J. Wampler for a high-resolution optical spectrum taken with the Lick Coude Echelle, G. Ferland for the most recent version of his ionization code CLOUDY, and Th. Kirsch for the ionization calculations. This work was supported by the Verbundforschung of the Bundesministerium für Forschung und Technologie.

Mechanical detection of magnetic resonance

D. Rugar*†, C. S. Yannoni* & J. A. Sidles‡

* IBM Research Division, Almaden Research Center, 650 Harry Road, San Jose, California 95120, USA

‡ University of Washington, Department of Orthopaedics, RK-10, Seattle, Washington 98195, USA

CONVENTIONAL techniques for measuring magnetic resonance involve the detection of electromagnetic signals induced in a coil or microwave cavity by the collective precession of magnetic moments (from nuclei or electrons) excited by an alternating magnetic field. In a different approach¹, isolated electron spins have been detected by scanning tunnelling microscopy, with the spin precession inducing a radiofrequency modulation in the tunnelling current. Here, we describe a new and extremely sensitive method of detection, the principles of which derive from magnetic force microscopy^{2–5} and a recent proposal^{6,7} by one of us (J.A.S.). We measure the small, oscillatory magnetic force (10^{-14} N) acting on a paramagnetic sample (a few grains of diphenylpicrylhydrazil, weighing <30 ng) which has been excited into magnetic resonance in the presence of an inhomogeneous magnetic field. This force is detected by optically sensing the angstrom-scale vibration of a micromechanical cantilever on which the sample is mounted. The sensitivity of this technique to the spatial distribution of the spins suggests that mechanical detection of magnetic resonance has the potential for imaging microscopic samples in three dimensions. So far, we have achieved a spatial resolution of 19 μm in one dimension.

The basic experimental principles of the force detection technique can be understood with the aid of Fig. 1a. A spin-containing sample is mounted on a micromechanical cantilever and placed close to a magnetized ferromagnetic particle. At the sample, the magnetic field emanating from the ferromagnetic particle is in the $\hat{z}$ direction. This field partially polarizes the spins in the sample, creating a small magnetization M_z . Because the magnetic field is spatially inhomogeneous, the sample experiences a force in the direction of the magnetic field gradient⁸. For the configuration in Fig. 1a, the force is in the $\hat{y}$ direction and has the magnitude $F_y = VM_z(\partial B_z/\partial y)$, where B_z

is the field at the sample and V is the sample volume. The force is converted into a mechanical deflection by the compliant cantilever spring on which the sample is mounted. To achieve the largest mechanical response, the force should oscillate at the mechanical resonance frequency of the cantilever, ω_c . When this is the case, the cantilever will vibrate with an amplitude $A = F_y Q/k$, where k is the cantilever spring constant and Q is the quality factor of the mechanical resonance.

To apply the magnetic force detection principle to the sensing of magnetic resonance, it was first necessary to devise a scheme that uses magnetic resonance effects to create the required oscillating force. One possible method would be to excite the spins in the sample with a radiofrequency (RF) magnetic field oscillating at the Larmor frequency $\omega_0 = \gamma B_z$, where γ is the gyromagnetic ratio ($\gamma = 1.76 \times 10^{11} \text{ s}^{-1} \text{ T}^{-1}$ for electrons). Excitation at this frequency creates a net torque on the spins, causing the axis of magnetization to tilt towards the x - y plane and begin precessing at frequency ω_0 about the z -axis⁹. In the presence of a magnetic field gradient, this precessing magnetization will create an oscillating magnetic force with frequency ω_0 . Unfortunately, ω_0 is typically in the megahertz frequency range and thus is not suitable for exciting vibrations in mechanical cantilevers, which typically have mechanical resonance frequencies in the kilohertz range.

To overcome this mismatch between mechanical and magnetic oscillation frequencies, we developed a magnetic resonance modulation technique that creates an oscillation of the longitudinal (z -axis) component of magnetization. To understand the technique, we first consider the behaviour of the longitudinal magnetization when the sample is subjected to a linearly polarized RF field, B_1 , directed at right angles to B_z . This case has been investigated theoretically¹⁰. The steady-state longitudinal magnetization is given by

$$M_z = (\chi_0/\mu_0) \left[B_z - \frac{\gamma^2 \omega^2 B_1^2 B_z \tau^4}{[1 + (\gamma B_z - \omega)^2 \tau^2][1 + (\gamma B_z + \omega)^2 \tau^2] + \frac{1}{2} \gamma^2 B_1^2 \tau^2 (1 + \omega^2 \tau^2 + \gamma^2 B_z^2 \tau^2)} \right] \quad (1)$$

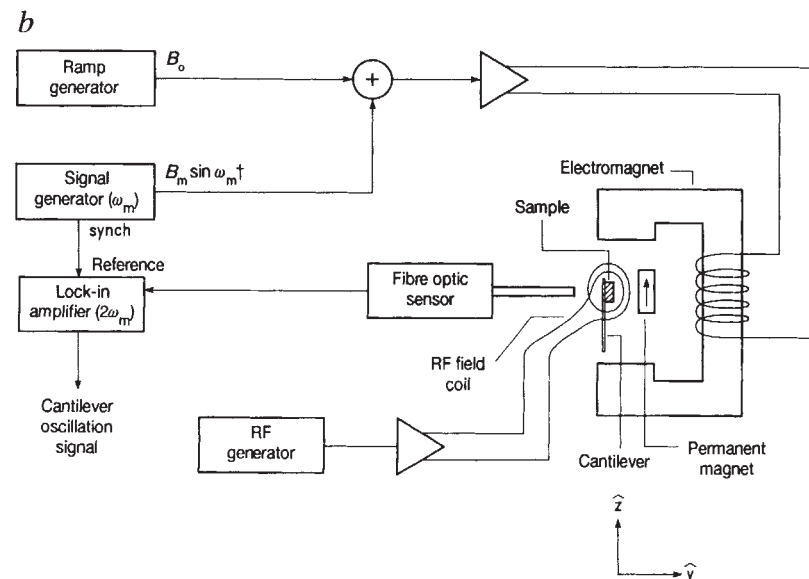
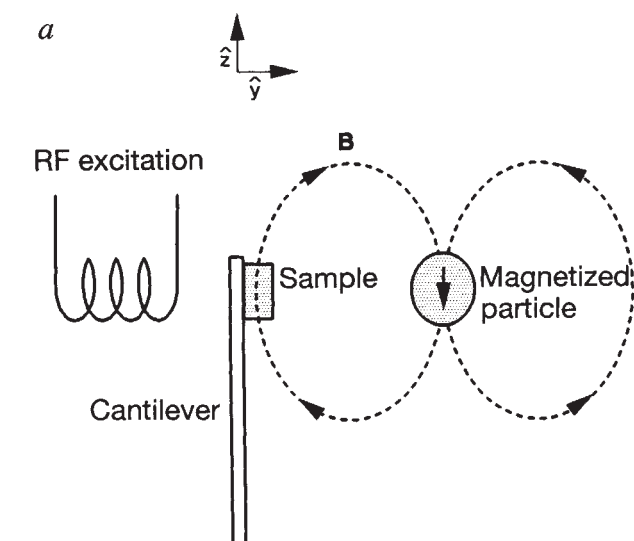
where χ_0 is the paramagnetic susceptibility in SI units, B_1 is the strength of the RF field, B_z is the strength of the polarizing magnetic field, ω is the angular frequency of B_1 , τ is the relaxation time and μ_0 is the permeability of vacuum. For diphenylpicrylhydrazil (DPPH), we use^{11,12} $\chi_0 = 2.5 \times 10^{-5}$ and $\tau = 6.2 \times 10^{-8} \text{ s}$. The SI value of susceptibility is obtained from

† To whom correspondence should be addressed.

the c.g.s. mass susceptibility¹² by multiplying by $4\pi\rho$, where ρ is the density of DPPH. The upper curve in Fig. 2 shows $M_z(B_z)$ for an applied RF field strength of $B_1 = 1$ mT. As can be seen in the figure, M_z is a linear function of B_z except in the vicinity of resonance. When the condition for magnetic resonance is met ($B_z = \pm \omega/\gamma$), the longitudinal magnetization is sharply suppressed.

Examination of equation (1) shows that there are at least three ways to create an oscillating M_z : modulating the amplitude of B_1 , modulating the frequency of B_1 , or modulating B_z . Whitfield and Redfield¹¹ demonstrated that amplitude modulation of B_1 could be used to cyclically suppress M_z . We tried this approach, but found that it created unacceptable levels of spurious vibration in the cantilever (probably due to eddy current and electrostatic force effects). Although we found that frequency modulation produced acceptable signals, most of our results were obtained by modulating B_z . In particular, we used a modulation of the form $B_z(t) = B_0 + B_m \sin(\omega_m t)$, where B_m is the modulation amplitude and ω_m is the modulation frequency. This modulation generates a time-varying $M_z(t)$ which, because of the nonlinear functional form of equation (1), has the harmonic expansion

$$M_z(t) = M_0 + M_1 \sin(\omega_m t) + M_2 \sin(2\omega_m t) + \text{higher harmonics} \quad (2)$$



For modulation of B_z at half the resonant frequency of the cantilever, $\omega_m = \omega_c/2$, the M_2 term resonantly excites the cantilever. This frequency-doubling technique was highly effective in reducing spurious (non-spin-resonance) excitation of the cantilever. For small values of B_m , M_2 is proportional to $\partial^2 M_z / \partial B_z^2$. Thus, M_2 is zero when $M_z(B_z)$ is a linear function and is non-zero in regions where $M_z(B_z)$ curves. This behaviour can be seen in the lower plot in Fig. 2, where M_2 is nearly zero except near resonance.

A block diagram of the overall experimental apparatus is shown in Fig. 1b. The sample consisted of a few grains of DPPH (grain size $\sim 10 \mu\text{m}$) which were attached with epoxy to the end of a thin-film silicon nitride cantilever. The $200 \mu\text{m}$ long cantilever was of the type commonly used for atomic force microscopy¹³ and had a spring constant of $k = 0.1 \text{ N m}^{-1}$. With the sample attached, the cantilever resonance frequency decreased from 20 kHz to 8 kHz, indicating that the total attached mass (DPPH plus epoxy) was $\sim 3 \times 10^{-8} \text{ g}$. We estimate that DPPH constituted less than half of the total attached mass. The mechanical resonance of the loaded cantilever showed a quality factor Q of 2,000 in vacuum. Vibration of the cantilever was sensed using a fibre-optic interferometer¹⁴ with sensitivity better than $10^{-3} \text{ \AA Hz}^{-1/2}$. The experiment was done at room temperature.

The ferromagnetic particle generating the inhomogeneous field consisted of a NdFeB permanent magnet. Two different particles were used: (1) a 3-mm-long, 3-mm-diameter cylindrical magnet which generated a field gradient of 10 T m^{-1} at the sample, and (2) a smaller magnet fragment roughly 1 mm in size, which generated a field gradient of 60 T m^{-1} . The inhomogeneous field from the permanent magnet was supplemented by the homogeneous field from a small electromagnet, which generated the required modulation field and provided a means to sweep the overall field strength. The electron spins in the sample were excited by a linearly polarized RF magnetic field (220–800 MHz) produced by a millimetre-size coil. The coil was impedance-matched to the 50Ω signal source using conventional resonant circuit techniques and driven with RF power of the order of 0.1–0.5 W.

Figure 3a shows the successful detection of electron spin resonance for the following experimental conditions: $\partial B_z / \partial y = 10 \text{ T m}^{-1}$, $\omega = 2\pi \times 220 \text{ MHz}$ and $B_m = 1 \text{ mT}$. The figure shows that peaks in the cantilever oscillation amplitude occur when the spin resonance condition is met ($B_z = \pm 7.9 \text{ mT}$). The maximum vibration amplitude is $\sim 2.8 \text{ \AA}$, corresponding to a peak force of $1.4 \times 10^{-14} \text{ N}$.

FIG. 1 a, Basic configuration used to detect magnetic resonance through magnetic force interactions. The spin-containing sample experiences a force due to the inhomogeneous magnetic field emanating from a nearby magnetic particle. Spins in the sample are excited by an applied RF field and a magnetic resonance modulation technique is used to modulate the magnetization of the sample. This causes the force on the sample to oscillate and the cantilever to vibrate. b, Block diagram of the overall apparatus. The vibration of the cantilever is at the second harmonic of the modulation frequency and is detected with a fibre-optic interferometer and a lock-in amplifier.

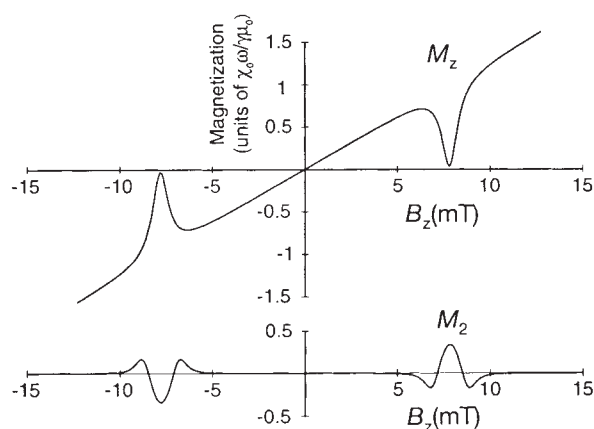


FIG. 2 The upper curve shows calculated magnetization M_z as a function of the applied field B_z for a DPPH sample subjected to a RF field of frequency 220 MHz and amplitude 1 mT. M_z is suppressed when the spin resonance condition is met ($B_z = \pm 7.9$ mT). The normalization factor, $\chi_0 \omega / \gamma \mu_0$, has the value $1.6 \times 10^{-1} \text{ A m}^{-1}$. The lower curve shows the strength of the second-harmonic component M_2 which results from modulating the polarizing field with an amplitude of $B_m = 1$ mT. The peak magnitude of M_2 is $\pm 0.33 \chi_0 \omega / \gamma \mu_0$, showing that second-harmonic conversion occurs with reasonable efficiency.

The peaks show the characteristics that one would expect of electron spin resonance: (1) they have the correct separation in terms of B_z ; (2) they were observed only when the RF excitation B_1 and the modulation B_m were present simultaneously; and (3) their positions shifted appropriately when the frequency of the RF excitation was changed. Furthermore, the experimental results are in reasonable agreement with theoretical expectations based on equations (1) and (2). Figure 3b shows a calculation of $M_2(B_z)$ for the experimental conditions described previously and assuming a point-like sample volume. Comparison of the two curves in Fig. 3 shows that the cantilever oscillation amplitude closely follows the theoretical behaviour of M_2 . This agreement is expected because the oscillatory magnetic force is proportional to M_2 .

The signal-to-noise ratio of the data improved with larger polarizing fields and larger field gradients. Results for $\partial B_z / \partial y = 60 \text{ T m}^{-1}$ and $\omega = 2\pi \times 800 \text{ MHz}$ are shown in Fig. 4a. At the higher field gradient, the spatial extent of the sample becomes important, resulting in a more complex peak structure. What was observed as a single positive-going peak for lower field gradient is now split into two distinct positive peaks separated by $\Delta B_z = 1.1 \text{ mT}$. The modelling results in Fig. 4b show that the

two peaks can be explained by assuming two concentrated regions of DPPH separated by $19 \mu\text{m}$. Inspection by optical microscopy confirmed that there were two distinct concentrations of DPPH on the cantilever (a result of the manual method used to glue the grains to the cantilever).

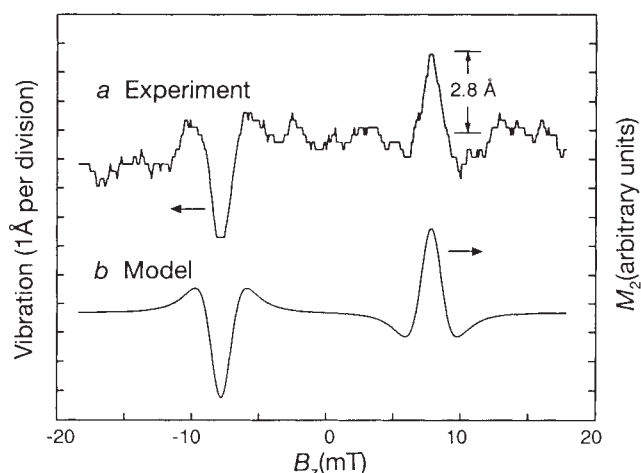
The high spatial resolution of the force detection technique is a natural consequence of the large field gradients we use to generate the magnetic force. Only those regions within a linewidth of resonance will generate an appreciable vibration signal. For a linewidth ΔB , the spatial resolution is $\Delta y = \Delta B / |\partial B_z / \partial y|$. Further improvement of resolution should be possible by using larger field gradients. For example, the gradient near the surface of a magnetized 100-nm-diameter iron particle (magnetization $\mu_0 M = 2.26 \text{ T}$) is $|\nabla B| = 8.5 \times 10^7 \text{ T m}^{-1}$. Assuming a linewidth of 1 mT, the spatial resolution would be $0.12 \text{ \AA}$. Because the magnetic force is proportional to the field gradient, a stronger field gradient increases the sensitivity of the detection technique, thus partially offsetting the reduction of signal caused by the smaller resonant sample volume.

In our experiment, the sensitivity of the mechanical detection technique is limited only by the thermal motion of the cantilever. The thermal noise force acting on the cantilever has a white frequency spectrum with the one-sided in-phase spectral density $S_f = 2kk_B T / Q\omega_c$, where k_B is Boltzmann's constant and T is temperature¹⁵⁻¹⁷. The minimum detectable force is $F_{\min} = (S_f \Delta \nu)^{1/2}$, where $\Delta \nu$ is the detection bandwidth in Hz. As the signal force magnitude is $M_2 V |\nabla B|$, the minimum detectable magnetic moment is¹⁸ $(M_2 V)_{\min} = (2kk_B T \Delta \nu / Q\omega_c)^{1/2} / |\nabla B|$. For our experimental conditions, $T = 300 \text{ K}$, $\Delta \nu = 0.1 \text{ Hz}$, $k = 0.1 \text{ N m}^{-1}$, $Q = 2,000$, $\omega_c = 2\pi \times 8 \text{ kHz}$ and $|\nabla B| = 60 \text{ T m}^{-1}$, we find $F_{\min} = 9 \times 10^{-16} \text{ N}$ and $(M_2 V)_{\min} = 1.5 \times 10^{-17} \text{ A m}^2$. This minimum detectable magnetic moment is equivalent to a $(3 \mu\text{m})^3$ volume of DPPH polarized in a 28.6-mT field (corresponding to 800 MHz spin resonance).

The sensitivity of the experiment could be greatly increased by using lower temperatures, higher gradients and more sensitive cantilevers. For the following parameters (which we believe are realizable), $T = 10 \text{ mK}$, $k = 0.01 \text{ N m}^{-1}$, $Q = 10^5$, $\omega_c = 10^6 \text{ s}^{-1}$, $\Delta \nu = 1 \text{ Hz}$ and $|\nabla B| = 10^8 \text{ T m}^{-1}$, the force sensitivity is $1.7 \times 10^{-19} \text{ N}$ and the magnetic moment sensitivity is $1.7 \times 10^{-27} \text{ A m}^2$. This suggests that detection of a single proton having moment $1.4 \times 10^{-26} \text{ A m}^2$ should be feasible in an analogous nuclear magnetic resonance (NMR) experiment. If such high sensitivity could be achieved, it might be possible to image directly the three-dimensional structure of complex molecules with single-atom resolution¹⁷.

Achieving this level of sensitivity and resolution will be a challenge for experimentalists. Among other things, it will require building an instrument compatible with cooling by a ^3He - ^4He dilution refrigerator. It also requires measuring

FIG. 3 a, Amplitude of cantilever vibration signal plotted as a function of the polarizing field for an applied modulation of $B_m = 1$ mT and RF excitation at 220 MHz. Vibration peaks with amplitude $\pm 2.8 \text{ \AA}$ occur at field values corresponding to electron spin resonance ($\pm 7.9 \text{ mT}$). Data shown is the average of 54 sweeps of the polarizing field, each sweep lasting 20 s. Lock-in amplifier time constant was 300 ms. b, Calculation of second-harmonic magnetization M_2 based on equation (1) and using known experimental parameters. The value of B_1 was not accurately known and was chosen to be 3.4 mT on the basis of the best fit with the data in a.



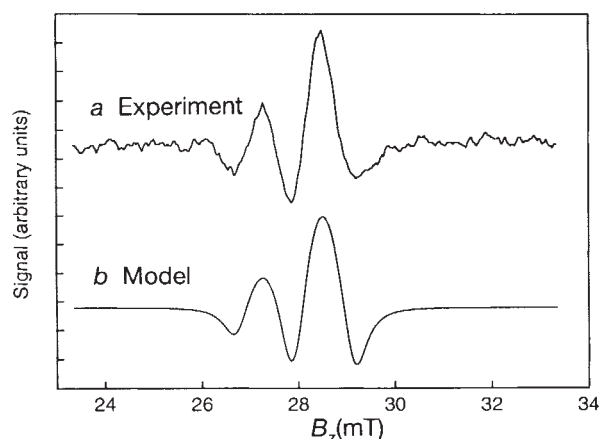


FIG. 4 *a*, Cantilever vibration signal measured using 800 MHz RF excitation and a field gradient of 60 T m^{-1} . Two positive-going resonance peaks separated by 1.1 mT are evident. Polarizing field modulation amplitude was $B_m = 0.7 \text{ mT}$. *b*, M_2 modelling result assuming two point particles with mass ratio 2.9:1 and separated by $\Delta y = 19 \mu\text{m}$. The values of B_z along the abscissa correspond to the field strength at the position of the larger particle. The fitted value of B_1 was 0.6 mT.

cantilever oscillation signals of $\sim 10^{-2} \text{ \AA}$, while simultaneously preventing spurious excitation of the cantilever by environmental noise or by back-action noise from the interferometer (photon pressure noise). In principle, these obstacles can be

overcome with careful experimental design. Scanning acoustic microscopes have already been operated successfully in a dilution refrigerator environment¹⁹, and a mechanical parametric amplifier has been developed that circumvents back-action noise effects²⁰. Although other difficulties will undoubtedly be encountered, we believe that the possibility of achieving atomic-scale NMR imaging would justify the effort. □

Received 6 July; accepted 23 October 1992.

- Manassen, Y., Hamers, R. J., Demuth, J. E. & Castellano, A. J. *Jr Phys. Rev. Lett.* **62**, 2531–2534 (1989).
- Martin, Y. & Wickramasinghe, H. K. *Appl. Phys. Lett.* **50**, 1455–1457 (1987).
- Saenz, J. J. *et al. J. appl. Phys.* **62**, 4293–4295 (1987).
- Rugar, D. *et al. J. appl. Phys.* **68**, 1169–1183 (1990).
- Grütter, P., Mamin, H. J. & Rugar, D. in *Scanning Tunneling Microscopy II* (eds Wiesendanger, R. & Güntherodt, H.-J.) 151–207 (Springer, Berlin, 1992).
- Sidles, J. A. *Appl. Phys. Lett.* **58**, 2854–2856 (1991).
- Sidles, J. A. *Phys. Rev. Lett.* **68**, 1124–1126 (1992).
- Jackson, J. D. *Classical Electrodynamics*, 185 (Wiley, New York, 1975).
- Abraham, A. *The Principles of Nuclear Magnetism*, 39–57 (Oxford Univ. Press, London, 1961).
- Garstens, M. A. & Kaplan, J. I. *Phys. Rev.* **99**, 459–463 (1955).
- Whitfield, G. & Redfield, A. G. *Phys. Rev.* **106**, 918–920 (1957).
- van Itterbeek, A. Labro, M. *Physica* **30**, 157–160 (1964).
- Albrecht, T. R., Akamine, S., Carver, T. E. & Quate, C. F. *J. Vac. Sci. Technol.* **A8**, 3386–3396 (1990).
- Rugar, D., Mamin, H. J. & Guethner, P. *Appl. Phys. Lett.* **55**, 2588–2590 (1989).
- Heer, C. V. *Statistical Mechanics, Kinetic Theory, and Stochastic Processes*, 431 (Academic, New York, 1972).
- Dürig, U., Züger, O. & Stalder, A. *J. appl. Phys.* **72**, 1778–1798 (1992).
- Sidles, J. A., Garbini, J. L. & Drobny, G. P. *Rev. Sci. Instrum.* **63**, 3881–3899 (1992).
- Gibson, G. A. & Schultz, S. *J. appl. Phys.* **69**, 5880–5882 (1991).
- Foster, J. S. & Rugar, D. *IEEE Trans. Sonics and Ultrasonics* **SU-32**, 139–151 (1985).
- Rugar, D. & Grütter, P. *Phys. Rev. Lett.* **67**, 699–702 (1989).

ACKNOWLEDGEMENTS. We thank T. Albrecht, P. Grütter, H. J. Mamin, P. Guethner and P. Wimmer for their contributions to the experimental apparatus. We also thank T. Albrecht, G. Drobny, J. Garbini and R. D. Kendrick for helpful discussions.

Superconductivity in barium fulleride

A. R. Kortan, N. Kopylov, S. Glarum, E. M. Gyorgy, A. P. Ramirez, R. M. Fleming, O. Zhou, F. A. Thiel, P. L. Trevor & R. C. Haddon

AT&T Bell Laboratories, Murray Hill, New Jersey 07974, USA

INTERCALATING solid C_{60} with dopant atoms yields materials with a remarkable range of properties¹. Alkali metal atoms, for example, readily form charge-transfer compounds^{2,3}, A_xC_{60} (where A is an alkali metal), which can be metallic, superconducting or insulating depending on the dopant concentration^{4–9}. In all cases, the superconducting phase has a face-centred cubic (f.c.c.) structure and stoichiometry A_3C_{60} , which suggests a common mechanism for superconductivity dependent, at least in part, on the external coordination number of the C_{60} molecules. More recently, it has been shown¹⁰ that the alkaline earth metal calcium can also be intercalated with fulleride to form a superconducting phase, again with a f.c.c.-derived structure, near a $\text{Ca}:\text{C}_{60}$ ratio of 5:1. Here we report the intercalation of fulleride with barium, in which a pure body-centred cubic phase with a lattice constant of 11.171 Å is realized near a stoichiometry of Ba_6C_{60} . This phase is also superconducting (with a transition temperature of 7 K), suggesting that the mechanism of superconductivity is related to an intrinsic property of the C_{60} molecules, rather than the external coordination number.

Sample preparation, synthesis and handling have been described previously¹⁰. Briefly, C_{60} is chromatographically purified and vacuum outgassed. High-purity barium metal is broken into a powder and mixed with C_{60} in a controlled-atmosphere glove box where the oxygen and water vapour levels are maintained below a few parts per million. Powder mixtures are then placed

in custom-designed high-purity tantalum cells which also allow compression into pellets. Over 40 samples with different compositions were prepared and loaded into quartz tubes without removing them from the tantalum cells. Other samples were prepared by loading the quartz tubes with unpressed powder mixtures. Although results for these samples were for the most part reproducible, we had to use tantalum cells for the reaction of barium to avoid contaminating the sample with unwanted reaction products. The quartz tubes were then connected to ultra-high-vacuum valves, taken out of the dry box and sealed under a vacuum of 10^{-6} torr. Heat treatments were carried out in tube furnaces at 550–800 °C, and for periods ranging from hours to weeks. Slow diffusion of barium in C_{60} creates serious problems with sample uniformity at temperatures lower than 600 °C. Highly uniform samples were prepared by subsequent anneals at 720 °C for 20 h, 550 °C for 60 h and 740 °C for 4 h. Following heat treatment, samples were removed from the tantalum cells in the glove box, loaded into 1-mm quartz X-ray capillaries and again sealed under high vacuum. We determined the appropriate reaction temperatures from X-ray diffraction experiments performed on mixed powders sealed in quartz capillaries and heated *in situ*.

Structural X-ray diffraction measurements were done with $\text{Mo K}\alpha$ radiation on a 12-kW rotating anode generator equipped with a triple axis goniometer. Flat ZYA graphite crystals were used to monochromatize and analyse the X-ray beam with a longitudinal resolution of $10^{-2} \text{ \AA}^{-1}$. A statistically averaged powder pattern was obtained by rocking the samples through 6° at each 2θ data point.

The crystal structure of the A_xC_{60} samples was studied for the composition range $2 < x < 8$. The diffraction pattern of Ba_6C_{60} samples fits well to a single body-centred cubic (b.c.c.) structure, similar to that found in A_6C_{60} (A is K, Rb or Cs) compounds. We have carried out a Rietveld refinement of the structure using the symmetry group $Im\bar{3}$. Refinement of 297

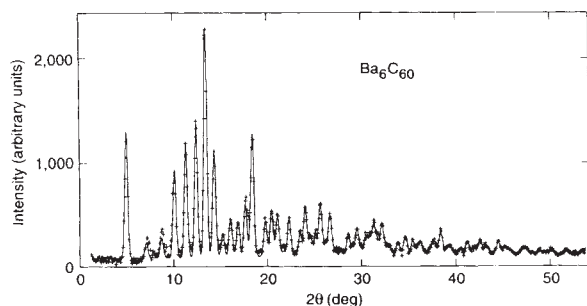


FIG. 1 Powder X-ray diffraction pattern of Ba_6C_{60} and the Rietveld refinement fit to a b.c.c. cell in the space symmetry $Im\bar{3}$.

peaks yielded a $11.171\text{-}\text{\AA}$ lattice constant, an intensity R factor of 9.6%, and the following coordinates: C1 at 0.0635, 0.0, 0.3087; C2 at 0.1262, 0.1029, 0.2693; C3 at 0.0635, 0.2057, 0.2301; and Ba at 0.0, 0.5, 0.2793. Mean-square thermal amplitudes of the isotropic Debye-Waller factors were refined to 0.016 and $0.023\text{ \AA}^2$ for C and Ba respectively, and all barium sites were found to be occupied. Figure 1 shows the powder X-ray diffraction pattern, and the refinement for the Ba_6C_{60} sample. A b.c.c. structure with an identical atomic decoration has been found for alkali-metal fullerenes¹¹. The measured lattice constant, $a_0 = 11.171\text{ \AA}$, however, is significantly smaller than measured values for K_6C_{60} , Rb_6C_{60} and Cs_6C_{60} ($11.39\text{ \AA}$, $11.52\text{ \AA}$ and $11.79\text{ \AA}$ respectively) and suggests a $9.67\text{-}\text{\AA}$ nearest-neighbour separation for the molecules. The size of a doubly ionized barium, $1.33\text{ \AA}$, placed at (0.22, 0.5, 0) requires a lattice constant of $11.744\text{ \AA}$ assuming spherically symmetric molecules. The relatively small lattice constant measured may indicate strong orbital overlap between the neighbouring molecules.

In the dilute regime, $x < 6$, X-ray diffraction peaks associated with a novel A15 Ba_3C_{60} phase coexists with the b.c.c. phase¹². Figure 2a shows the relative ratios of the strongest diffraction peaks of these phases, the 210 and 321 peaks of the A15 and the b.c.c. phases respectively. The volume fraction of the A15 phase relative to the b.c.c. phase decreased rapidly with increasing barium concentration and disappeared completely at $x = 6$.

The number of interstitial sites increases from three per C_{60} in a face-centred cubic (f.c.c.) structure to six in a b.c.c. structure.

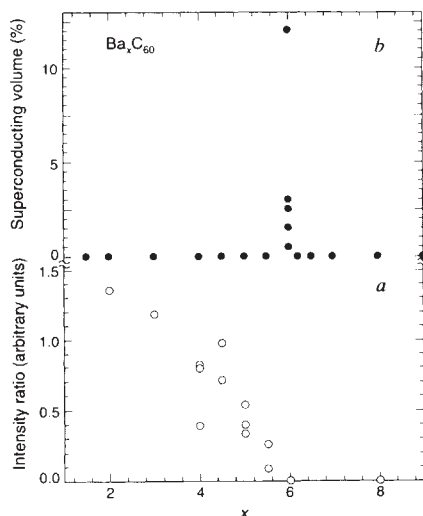


FIG. 2 Variations of the strongest diffraction peak ratios of the A15 to that of the b.c.c. phase (a), and percentage superconducting volume of the samples (b), with barium concentration.

Additionally, the distinction between octahedral and tetrahedral sites is removed in the b.c.c. structure, all interstitial sites becoming equivalent with distorted tetrahedral symmetry. The b.c.c. lattice constant of samples with nominal compositions $2 < x < 8$ was determined from fits to the diffraction peaks, and did not change with barium concentration; the value is $a = 11.171 \pm 0.015\text{ \AA}$, within the experimental resolution. Qualitatively, relative intensities of the b.c.c. diffraction pattern also did not change with x , suggesting that the b.c.c. phase formed in small grains and its volume fraction grows with x . These results seemed to be largely independent of the temperature treatment of the samples, indicating that kinetic limitations were minimal and samples were close to thermodynamic equilibrium. The lattice constant remained unchanged beyond $x = 6$, as 6 is the maximum number of barium atoms that the lattice can accommodate. In this range, extra diffraction peaks appeared, indicating the formation of other phases.

Superconductivity was observed only in samples within a very narrow range of compositions about Ba_6C_{60} (Fig. 2b). The superconducting volume fraction peaked sharply near $x = 6$ and had a maximum value of 12% for our best sample, as determined by magnetic susceptibility measurements. X-ray diffraction

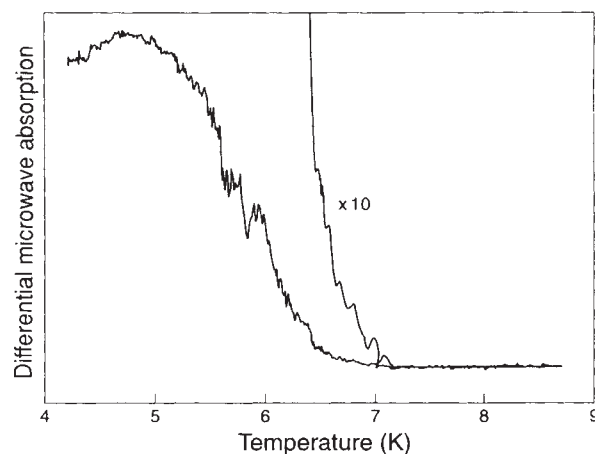


FIG. 3 Microwave loss measurements for Ba_6C_{60} in a static field of 100 Oe.

measurements (Fig. 1) revealed no evidence for a phase other than the b.c.c. at $x = 6$. The fact that the b.c.c. phase occurs over a wide range of stoichiometries, whereas superconductivity appears only at $x = 6$, may be explained by partial occupancy of the Ba sites for $x < 6$. An occupancy slightly less than one may result in intensity changes too small to be noticed. We estimate an upper bound for variations in occupancy to be 0.05 in changing composition from $x = 4$ to $x = 6$. The appreciable superconducting volume fraction and the lack of impurity peaks in Fig. 1 seem to rule out a third undetected phase as the source of superconductivity.

A variety of consistently weak signals were seen in electron spin resonance spectra. These varied greatly from sample to sample and, apart from a narrow, $g = 2.0022$ resonance attributed to thermal decomposition of C_{60} , could not be identified. No signal could be associated with the superconducting phase, as was also the case for Ca doping. The absence of an observable electron spin resonance signal from the conduction electrons may indicate spin-orbit relaxation by the metal due to partial charge transfer. Figure 3 shows a microwave absorption measurement for an $x = 6$ sample, with an onset at 7 K, in a field of 100 Oe at 9 GHz. The broad region between onset and the differential loss maximum indicates weak forces inhibiting vortex motion and is characteristic of less-well-ordered C_{60} films doped with alkali metals¹³.

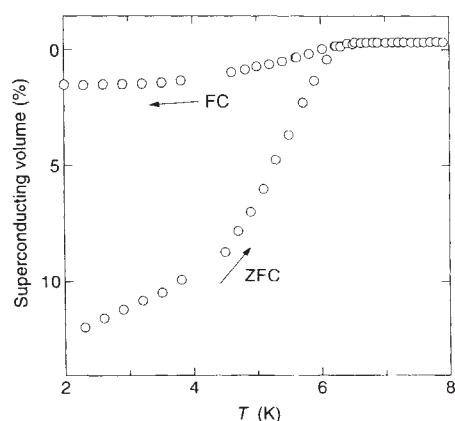


FIG. 4 Magnetic susceptibility measurements of the Ba_6C_{60} sample. FC, field cooled; ZFC, zero field cooled. In the zero-field-cooled case we find a diamagnetic volume fraction of 12%.

A SQUID magnetometer was used to measure the temperature dependence of the d.c. magnetization of the samples. Figure 4 shows these measurements for the sample shown in Fig. 3. The sample was first cooled at 2.2 K in zero field and then gradually warmed in a magnetic field at 5 Oe. This applied field was excluded to 6 K. At 2.2 K, a diamagnetic state with 12% shielding was measured. This value should be taken as a lower bound for the shielding fraction, because of finite size effects in granular superconductors. A magnetic flux expulsion in the superconducting state or the Meissner effect is observed when the sample was cooled in a field of 2 Oe. These measurements find a Meissner fraction of 2%.

In alkali-metal-intercalated fullerenes, the superconducting phase is f.c.c. A_3C_{60} with dopant atoms occupying the interstitial sites. Dopant atoms donate three electrons per C_{60} , half filling the t_{1u} -derived band. The size of the unit cell and the superconducting transition temperature increases with increasing cation size. Moving the fullerenes further apart sharpens the density of states at the Fermi level¹⁴⁻¹⁶ and raises the transition temperature. For divalent alkali or rare earth metals, the nature of the charge transfer is more complex. For nearly complete charge transfer, the C_{60} t_{1g} -derived band may become populated¹⁵. Recent electronic structure calculations¹⁷ for calcium-doped C_{60} find that for Ca_5C_{60} the t_{1u} is completely occupied and the carriers are in the t_{1g} band. Strong supporting evidence for the participation of this new band is also found in photoemission experiments (J. H. Weaver, personal communication). This is in contrast to the barium-intercalated graphite case, where X-ray photoemission and c-axis compressibility measurements¹⁸ find barium atoms donating only one of their two valence electrons to the graphite conduction band. This new superconducting phase, with six barium atoms filling all interstitial sites, is unlike all other known superconducting fullerene compounds as it possesses a b.c.c. rather than an f.c.c. structure. For the divalent Ca_5C_{60} superconducting phase, the observed reduction in symmetry from f.c.c. to a simple cubic structure¹⁰ is thought to arise from a rearrangement of the calcium atoms to symmetrically inequivalent positions within the unit cell: the C_{60} molecules themselves retain the f.c.c. structure found for the alkali metal superconducting fullerenes. A b.c.c. superconductor phase may indicate that the superconductivity mechanism is independent of the coordination number, but more related to an internal property of the molecule itself, like intramolecular phonon modes^{16,19}. As we have pointed out for the Ca_5C_{60} case, the direct products $t_{1u} \times t_{1u}$ and $t_{1g} \times t_{1g}$ give rise to identical representations, and thus the same C_{60} intramolecular vibrations^{16,19} could supply the electron-phonon coupling for superconductivity in both alkali and alkali-earth metal-doped compounds.

A primarily phonon-mediated mechanism for superconductivity in alkali-metal intercalated C_{60} has been found in recent isotope substitution experiments²⁰. □

Received 17 June; accepted 12 November 1992.

1. Fullerenes (eds Hammond, G. S. & Kuck, V. J.) (American Chemical Society, 1992).
2. Haddon, R. C. *et al.* *Nature* **350**, 320-322 (1991).
3. Stephens, P. W. *et al.* *Nature* **351**, 632 (1991).
4. Hebard, A. F. *et al.* *Nature* **350**, 600-601 (1991).
5. Holczer, K. *et al.* *Science* **252**, 1154-1157 (1991).
6. Rosseinsky, M. J. *et al.* *Phys. Rev. Lett.* **66**, 2830-2832 (1991).
7. Chen, C.-C., Kelty, S. P. & Lieber, C. M. *Science* **253**, 886 (1991).
8. Fleming, R. M. *et al.* *Nature* **352**, 701 (1991).
9. Zhou, O. *et al.* *Nature* **351**, 462 (1991).
10. Kortan, A. R. *et al.* *Nature* **355**, 529 (1992).
11. Fischer, J. E., Heiney, P. A., Luzzi, D. E. & Cox, D. E. in Fullerenes (eds Hammond, G. S. & Kuck, V. J.) (American Chemical Society, 1992).
12. Kortan, A. R. *et al.* preprint, AT&T Bell Labs; *Phys. Rev. Lett.* (submitted).
13. Glarum, S. H., Duclos, S. J. & Haddon, R. C. *J. Am. chem. Soc.* **114**, 1446 (1992).
14. Fleming, R. M. *et al.* *Nature* **352**, 787 (1991).
15. Haddon, R. C. *et al.* *ACS Symp. Series No. 481* (in the press).
16. Schluter, M., Lannoo, M., Needels, M., Baraff, G. A. & Tomanek, D. *Phys. Rev. Lett.* **68**, 526 (1992).
17. Saito, S. & Oshiyama, A. *Solid State Commun.* **83**, 107 (1992).
18. Fischer, J. E., Kim, H. J. & Cajipe, V. B. *Phys. Rev. B* **36**, 4449 (1987).
19. Varma, C. M., Zaanen, J. & Raghavachari, K. *Science* **254**, 989-992 (1991).
20. Ramirez, A. P. *et al.* *Phys. Rev. Lett.* **68**, 1058 (1992).

ACKNOWLEDGEMENTS. We thank T. Siegrist, D. W. Murphy, M. A. Schluter, A. F. Hebard, M. J. Rosseinsky and J. Tully for discussions.

Intercalation of sodium hetero-clusters into the C_{60} lattice

T. Yildirim*, O. Zhou*, J. E. Fischer*, N. Bykovetz*, R. A. Strongin*, M. A. Cichy*, A. B. Smith III*, C. L. Lin† & R. Jelinek‡

Laboratory for Research on the Structure of Matter, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

† Department of Physics, Temple University, Philadelphia, Pennsylvania 19122, USA

‡ Materials Sciences Division, Lawrence Berkeley Laboratories and Department of Chemistry, University of California, Berkeley, California 94720, USA

INTERCALATION of sodium into C_{60} has been shown¹ to yield a range of compounds Na_xC_{60} ($2 < x < 6$). Unlike the other alkali-metal-doped compounds, there is no evidence of superconductivity within this range of doping. Furthermore, Na_6C_{60} retains the face-centred-cubic (f.c.c.) structure of undoped C_{60} whereas the analogous $x = 6$ phases with K, Rb or Cs are body-centred cubic (b.c.c.)². Here we report the preparation of sodium-doped C_{60} with x up to about 10; the structure remains f.c.c. throughout. Rietveld refinement of X-ray diffraction data yields a structure with ideal composition $\text{Na}_{11}\text{C}_{60}$, in which the octahedral interstitial site contains a nine-atom body-centred cluster of sodium atoms while the tetrahedral sites are singly occupied. Clusters of four to nine sodium atoms in the octahedral site provide sufficient 'chemical pressure' to prevent the low-temperature lattice distortion that occurs when this site is singly occupied, as in Na_3C_{60} (ref. 1). This distortion is thought to suppress superconductivity in the latter compound. Thus we suggest that new sodium-doped superconducting phases may be found in the range $4 < x < 11$.

Rapid progress has been made in understanding the systematics of heavy alkali intercalation into the interstices of crystalline C_{60} fullerene^{3,4}. Attention has mainly been focused on $[\text{M}_y\text{M}'_{1-y}]_x\text{C}_{60}$ binary and ternary face-centred-cubic (f.c.c.) compounds with $x = 3$ and where M and M' represent K, Rb or Cs. Lattice constants are generally slightly larger than in pure

* T.Y. is also at the Physics Department, University of Pennsylvania, O.Z., J.E.F. and N.B. are also at the Department of Materials Science and Engineering, University of Pennsylvania; present address for O.Z. is AT&T Bell Laboratories, Murray Hill, New Jersey 07974, USA. R.A.S., M.A.C. and A.B.S. are also at the Chemistry Department, University of Pennsylvania. Correspondence should be sent to J.E.F.

C_{60} and seem to be controlled by 'chemical pressure' of the alkali ion in the tetrahedral interstitial sites (two per molecule). X-ray refinements give large thermal factors for the ion in the considerably larger octahedral site⁵ (one per molecule), possibly indicating static disorder due to the mismatch between ionic and site volumes¹. An empirical correlation between superconducting transition temperature T_c and lattice constant has been established, which in the context of BCS theory implies that T_c is mainly controlled by the width of a half-filled band derived from the t_{1u} lowest unoccupied molecular orbital (LUMO)^{6,7}. Doping with more than three M atoms per C_{60} induces structural transitions to body-centred tetragonal (b.c.t.) at $x = 4$ (ref. 8) and to b.c.c. at the maximum concentration $x = 6$ (ref. 2), because the f.c.c. lattice cannot expand enough to accommodate more than one K, Rb or Cs per f.c.c. interstitial site of either type.

Sodium-containing binaries and ternaries show very different behaviour^{1,9}. Na_3C_{60} does not superconduct, despite the fact that the 300 K f.c.c. lattice constant implies a $T_c > 10$ K, by analogy with the other alkali-metal-doped materials. Similarly, f.c.c. $Na_2M'C_{60}$ ternary phases exhibit only small diamagnetic shielding fractions with anomalously low values of T_c (refs 4, 9). Most surprising of all, Na_6C_{60} remains f.c.c., with tetrahedral sites singly occupied but the octahedral site now containing a four-atom Na tetrahedron¹. Structural analysis of the compound described here shows that the octahedral site can contain up to nine Na atoms, leading to an ideal maximum composition $Na_{11}C_{60}$. Some features of this structure and its temperature dependence provide clues about the possible occurrence of superconductivity in this sub-family of fullerene intercalation compounds.

Reactions were carried out in evacuated 0.5-cm-diameter oxygen-free high-conductivity (OFHC) copper tubes, which are

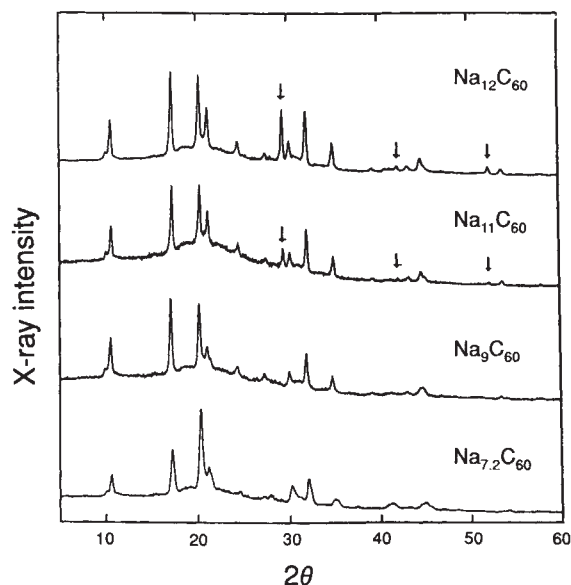


FIG. 1 X-ray powder diffraction patterns of four Na_xC_{60} samples. Reactions were carried out in evacuated Cu tubes (10^{-4} torr), double-sealed in evacuated pyrex to avoid oxidizing the outer Cu surface. A typical heating schedule consisted of 6 hours at 300 °C followed by 1, 3, 5 and 3 days at 500 °C, 450 °C, 400 °C and 350 °C, which includes higher temperatures and longer times than used by Rosseinsky *et al.*¹. We suspect however that the more important difference is our use of evacuated tubes, and that the previous limiting composition Na_6C_{60} is probably a kinetic effect. The top three profiles result from stoichiometric reactions with nominal x values of 12, 11 and 9. Peaks indicated by arrows for $x = 12$ and 11 are due to unreacted sodium, absent for $x = 9$. The bottom profile is obtained by distilling Na out of the $x = 12$ sample at 350 °C for 4 hours, followed by a 4-day anneal at the same temperature ($x = 7.2$ by titration).

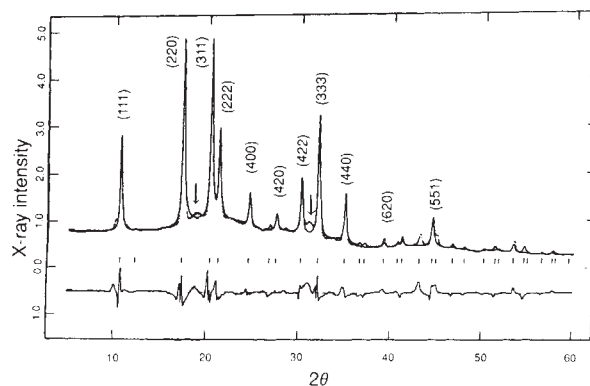


FIG. 2 X-ray powder diffraction pattern (dots) of the saturated compound $Na_{9.7}C_{60}$, a Rietveld profile refinement in space group $fm\bar{3}$, $R_{wp} = 9.9\%$ (solid curve), and a difference plot (bottom). Na atoms are located at the tetrahedral and octahedral sites, and at the general point (x, x, x) with fractional occupancies 0.88, 0.40 and 0.89 respectively. The refined value of x is 0.39 and $a = 14.59$ Å, resulting in a nearest Na-Na distance of 2.78 Å along the body diagonal. Data were collected on an INEL diffractometer equipped with a 120° detector, flat graphite monochromator and Cu radiation. Broad features indicated by arrows are attributed to a highly faulted h.c.p. minority phase.

readily vacuum sealed by swaging and offer good resistance to molten Na up to 600 °C (ref. 10). Titrations of several Na-saturated samples gave $x = 9.7 \pm 0.4$. To verify this independently as the maximum doping level, we prepared several samples using predetermined amounts of C_{60} and Na to make $x = 12$, 11 or 9, having first verified that Na loss by reaction with Cu was negligible¹⁰. Figure 1 shows X-ray profiles obtained from these reactions (top) and from a sample exposed to molten Na which gave $x = 7.2$ by titration after distilling off excess Na and annealing (bottom). The profiles for $x = 12$ and $x = 11$ are identical, both containing weak peaks from free Na indicating a maximum Na concentration slightly less than 11. The $x = 9$ sample shows no free Na peaks and the relative intensities differ significantly from those obtained with $x = 12$ and 11; this means that the maximum x exceeds 9. All three profiles show relative intensities substantially different from those of Na_6C_{60} (ref. 1). Using the intensities of free Na peaks scaled to a common peak in the samples with $x = 11$ and 12, it is possible to estimate the actual Na concentration in the saturated phase. We thus find $x = 9.8 \pm 0.5$, in agreement with the titration results.

Figure 2 shows the X-ray profile from a sample reacted with excess Na but without liquid contact ($Na_{9.7}C_{60}$ by titration). Dots are the experimental data, the solid curve is the Rietveld fit and a difference plot is shown below. Most reflections can be indexed on a f.c.c. lattice with $a = 14.59$ Å, significantly expanded relative to the 14.38 Å value for Na_6C_{60} (ref. 1). Other important differences are evident. In the saturated compound the (220) and (311) intensities are similar whereas for $x = 6$ the (311) is 3 times stronger. We also find observable intensities at the (400), (420), (620) and (444) positions, all of which are very weak or absent for $x = 6$. All samples contain a minority hexagonal-close-packed (h.c.p.) constituent. The Rietveld refinement is based on a model derived from the one proposed for Na_6C_{60} (ref. 1) except with 11 Na sites per C_{60} (space group $fm\bar{3}$): the two tetrahedral sites, the high-symmetry octahedral position $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$ and the eight corners of a cube centred at $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$. Figure 3 gives a schematic representation. The octahedral 'site' contains a body-centred Na_3 cluster; the cube edge refines to 3.21 Å, significantly less than for b.c.c. Na metal ($a = 4.235$ Å) but larger than a hypothetical cube of Na^+ ions in contact (2.34 Å) which would certainly be electrostatically unstable. Refined site occupancies are 0.88, 0.40 and 0.89 for tetrahedral, octahedral and cube-corner sites respectively, yielding $x = 9.28$,

consistent with the other determinations discussed above. Inclusion of the octahedral (body-centre) Na improves the refinement. The model is supported by static ^{23}Na NMR (Fig. 4), which features three peaks at 172 p.p.m., 3 p.p.m. and -121 p.p.m. with fitted lorentzian intensity ratios 1.0:0.26:0.17 respectively. It is tempting to assign these to cluster corner, isolated tetrahedral and cluster centre Na atoms, respectively, which for the ideal stoichiometry would be present in the ratio 1.0:0.25:0.12. On the other hand, one would have expected *a priori* that the corner Na atoms would be more ionic, or less metallic, than the centre one. Further work is in progress to identify the three resonances with specific Na sites. No narrowing was observed when we tried 5-kHz magic-angle spinning or double rotation, allowing us to rule out anisotropic interactions at the Na nuclei as an important factor in the linewidth.

It is clear from the difference plot in Fig. 2 that the agreement would be improved by a two-phase analysis including the minority h.c.p. component. The C_{60} molecules are orientationally ordered, as in b.c.c. M_6C_{60} ($\text{M} = \text{K}, \text{Rb}$ or Cs), as the $f\bar{m}\bar{3}$ space group correctly reproduces the observed (420) intensity whereas $f\bar{m}\bar{3}m$ does not. Other degrees of freedom to be explored include the precise geometry of the Na_9 clusters, which will probably be different in the oxidizing C_{60} environment from the geometry of free neutral clusters^{11,12}. A general distortion away from the assumed cubic symmetry would make x, y and z inequivalent in the crystal environment. Therefore, any relaxation of the cluster from local cubic symmetry implies that the clusters must be orientationally disordered to preserve the cubic lattice symmetry. If such disorder exists, it is most likely to be static on the X-ray timescale even at 300 K; an X-ray profile taken at 81 K gives no evidence of the symmetry-lowering that would be imposed by orientational order of the Na clusters. Preliminary theoretical work indicates that a partially ionized, unrelaxed octahedral Na_9 b.c.c. cube results in a stable lattice (J. L. Martins, personal communication).

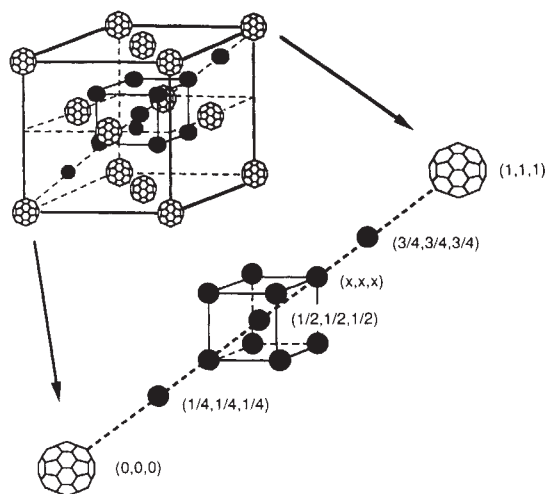


FIG. 3 Schematic representation of the structure derived from Fig. 2. For clarity, isolated tetrahedral Na atoms are shown along only one of the body diagonals, and only the cluster centred at (0.5, 0.5, 0.5) is shown. The C_{60} molecules are orientationally ordered. The fit was improved only slightly by allowing orientationally disordered distortions of the octahedral cluster. In the first, suggested by Fig. 4b of ref. 12, one of the cube faces is rotated by 45° to optimize electrostatic interactions in the neutral cluster. The second, prompted by Fig. 1 of ref. 11, consists of two nested tetrahedra with local T_d symmetry, similar to the geometry proposed for Na_6C_{60} . The small differences among the three models can be attributed to the observation that the average radial charge density is not very sensitive to the exact geometry¹². In fact all the calculations predict large charge density at the cluster centre, justifying the need for a partially occupied Na site at (0.5, 0.5, 0.5) in our model.

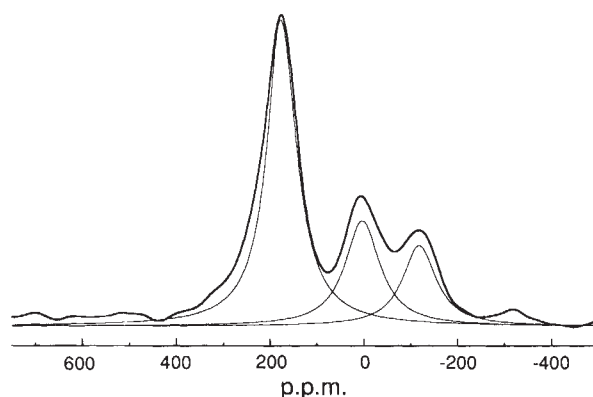


FIG. 4 Static ^{23}Na NMR spectrum of $\text{Na}_{9.7}\text{C}_{60}$ at 300 K. We used a Chemagetics CMX-500 spectrometer and 11.7 T magnetic field. Short 30° pulses were used with a 200 ms delay between acquisitions (50,000 scans). Data were zero-filled to 2 K points with application of 3,000 Hz Lorentzian apodization. The external reference was 0.1 M NaCl solution. The three-line spectrum is consistent with the model shown in Fig. 3.

Direct-current magnetization (SQUID) measurements show that the Na-saturated phase does not superconduct above 2 K, in common with $x = 2, 3$ and 6 phases¹. Milder reaction conditions, or distilling Na out of the saturated compound, both yield single-phase f.c.c. products with intermediate lattice constants and partial occupancies of all three types of sites. For instance, the bottom profile in Fig. 1 is from a sample prepared by distilling Na out of nominal $\text{Na}_{12}\text{C}_{60}$ at 350°C for 4 hours, followed by a 4-day anneal at the same temperature. Titration of this sample gave $x = 7.2 \pm 0.4$. Our results are generally consistent with solid-solution behaviour in the range $3 < x < 11$.

It seems likely that the absence of superconductivity in Na_3C_{60} is related to the low-temperature instability of the f.c.c. lattice. In the f.c.c. high-temperature phase, Na_3C_{60} almost certainly has an electronic spectrum similar to those of the superconductors, but a single octahedral Na stabilizes the f.c.c. lattice only if it undergoes random thermal motion of sufficient amplitude. A similar low-temperature distortion is observed in M_1C_{60} phases with $\text{M} = \text{K}, \text{Rb}$ or Cs ; these also contain a single M somewhere in the octahedral site but the tetrahedral sites are unoccupied (Q.Z. *et al.*, unpublished data). The distortion in both these compounds can be envisaged either as a consequence of freezing out thermal (dynamic) disorder, or as an Ising-like transition driven by mixing entropy of partially occupied symmetry-inequivalent sub-sites. In contrast, the 4- or 8(9)-atom octahedral cluster in Na_6C_{60} and the saturated phase respectively provide sufficient 'chemical pressure' to maintain the octahedral site volume without thermal fluctuations, and thus stabilize the f.c.c. structure at low temperature. Here the absence of superconductivity can be attributed to low values of $N(E_F)$, the density of states at the Fermi energy, with $x = 6$ and 10. It therefore seems possible that superconductivity will occur at some x value in the solid-solution range for which the f.c.c. structure is stable at low temperature, namely $4 < x < 11$. We would expect similar phenomena to occur in Li_xC_{60} , for which there is a saturation doping level $x \approx 11$, and a solid-solution range $3 < x < 11$ has also been reported¹³. □

Received 22 October; accepted 13 November 1992.

1. Rosseinsky, M. J. *et al.* *Nature* **356**, 416–418 (1992).
2. Zhou, O. *et al.* *Nature* **351**, 462–464 (1991).
3. Zhu, Q. *et al.* *Science* **252**, 545–548 (1991).
4. Murphy, D. W. *et al.* *J. Phys. Chem. Solids* **53**, 1321 (1992).
5. Stephens, P. W. *et al.* *Nature* **351**, 632–634 (1991).
6. Fleming, R. M. *et al.* *Nature* **352**, 787–788 (1991).
7. Zhou, O. *et al.* *Science* **255**, 833–835 (1992).
8. Fleming, R. M. *et al.* *Nature* **352**, 701–703 (1991).

9. Tanigaki, K. *et al.* *Nature* **356**, 419–421 (1992).
10. *Liquid-Metals Handbook*, 2nd. Edn. (U.S. Govt. Printing Office, Washington DC, 1954).
11. Koutecky, V. B., Fantucci, P. & Koutecky, J. *Phys. Rev.* **B37**, 4369–4374 (1988).
12. Rothlisberger, U. & Andreoni, W. *J. Chem. Phys.* **94**, 8129–8151 (1991).
13. Chabre, Y. *et al.* *J. Am. chem. Soc.* **114**, 764–765 (1992).

ACKNOWLEDGEMENTS. We thank A. Pines and Q. Zhu for discussions. This work was supported by the NSF Materials Research Laboratory Program and the Department of Energy.

Removal of sulphur from the marine boundary layer by ozone oxidation in sea-salt aerosols

H. Sievering^{*†}, J. Boatman[†], E. Gorman^{*}, Y. Kim[†],
L. Anderson^{*}, G. Ennis^{*}, M. Luria[‡] & S. Pandis[§]

^{*} Center for Environmental Sciences and Department of Physics, Box 136, University of Colorado-Denver, PO Box 173364, Denver, Colorado 80217, USA

[†] Aerosol Research Section, National Oceanic and Atmospheric Administration, Boulder, Colorado 80303, USA

[‡] Environmental Sciences Division, The Hebrew University, Jerusalem, Israel

[§] Department of Chemical Engineering, California Institute of Technology, Pasadena, California, USA

THE oxidation of sulphur dioxide to sulphate in the marine boundary layer (MBL) is an important pathway in the global sulphur cycle. Oxidation by ozone in the aqueous phase is an important process in cloud droplets¹ but has not generally been thought to be significant in the clear air of the MBL. Yet the lower part of the MBL contains abundant sea-salt aerosol particles, which are largely water of sufficiently high pH (ref. 2) to support ozone oxidation of SO₂ to sulphate. We have argued previously³ that 5–25% of the total non-sea-salt sulphate (n.s.s. SO₄²⁻) observed in the MBL may be formed by this mechanism; here we assess its contribution to the cycling of sulphur in (and particularly its removal from) the MBL. We show that, owing to the effects of mass transfer, the n.s.s. SO₄²⁻ so generated will be predominantly associated with particles of 2–9 µm diameter, and will accordingly dry-deposit at a rapid rate. Because part of the dimethyl sulphide (DMS) emitted by marine organisms is converted to SO₂ in the MBL, this additional removal pathway for sulphur may markedly reduce the proposed feedback⁴ between greenhouse warming, oceanic DMS emissions and sulphate haze albedo.

Careful observations of the size distribution of n.s.s. SO₄²⁻ in the MBL have shown that a substantial fraction of it is found in large, predominantly sea-salt particles^{5–7}. These large particles have high dry deposition rates, and removal of sulphur in this way to the ocean's surface must be assessed in MBL sulphur budgets. Depending on the sampling efficiency of sea-salt particles and meteorological conditions (locally or upwind), the observed n.s.s. SO₄²⁻ in the sea-salt particles may vary from an indeterminable amount to 5 nmol m⁻³ or more⁵. In recent sampling at Barbados⁸, more than 15% of the total observed n.s.s. SO₄²⁻ was found in particles more than 1.2 µm in diameter. The Coordinated Air-Sea Experiment (CASE) of the Global Change Expedition^{7,8}, operating in the Bermuda area in summer 1988, found that 16–18% of the observed n.s.s. SO₄²⁻ from shipboard sampling (>20% when considering six-stage impactor data; A. Pszenny, personal communication), and a similar or higher percentage from aircraft sampling, occurred in particles of >1 µm diameter, despite a small sea-salt mass under light wind-speed conditions^{7,9}. During sampling at higher wind speeds in the North Atlantic⁵ as much as half of the observed n.s.s. SO₄²⁻ was in particles of >0.9 µm diameter.

Modelling of the MBL sulphur cycle during Bermuda-area CASE¹⁰ showed that 50–70% of SO₂ oxidation must have been by heterogeneous conversion, and that a n.s.s. SO₄²⁻ deposition

velocity of nearly 1 cm s⁻¹ had to be invoked to obtain a balance between sulphur sources and sinks. This large deposition velocity supports the observation that 15–20% of total n.s.s. SO₄²⁻ was in large sea-salt particles. The model outcome that most SO₂ oxidation occurred by the heterogeneous pathway is especially interesting as cloud processing of SO₂ was at a minimum during this drought period at Bermuda producing <5% of the observed n.s.s. SO₄²⁻ in the lowest 15 to 150 m of the MBL³. Furthermore, heterogeneous SO₂ cloud conversion in nonprecipitating clouds produces no particles¹¹ more than 1 µm in diameter. Homogeneous SO₂ conversion certainly cannot explain the large-particle n.s.s. SO₄²⁻. The combination of cloud processing of SO₂ (including collision-coalescence within clouds) and coagulation of small-particle n.s.s. SO₄²⁻ cannot explain more than 1% of the large-particle n.s.s. SO₄²⁻ found by the Bermuda-area CASE³.

A heterogeneous SO₂ conversion mechanism^{2,3}, not previously considered in sulphur mass-balance estimates for the MBL, can explain the observed large-particle n.s.s. SO₄²⁻. This mechanism relies on oxidation of aqueous-phase sulphur, S(IV), in the high-pH water of sea-salt aerosol particles. The pH of sea-salt aerosol water (SSAW) may be expected to fall in the range 6.5–7.8, with pH 7 the most likely². The rate expressions for aqueous oxidation of S(IV), valid up to pH 7 and at ambient MBL temperatures (see for example Seinfeld¹), are applicable here. At pH 7, oxidation by ozone (O₃) produces the most rapid conversion of S(IV) to n.s.s. SO₄²⁻ in the SSAW. Iron and other catalysts may also contribute; however, we will consider only O₃ oxidation, noting that other means of oxidation will enhance the actual oxidation rate.

The rate of SO₂ and O₃ mass transfer into the SSAW must be assessed to identify aqueous-phase S(IV) and O₃ concentrations as input to the S(IV) oxidation-rate expressions above. Aqueous-phase diffusion of O₃, SO₂ gas diffusion and interfacial mass transfer are rate-limiting steps¹² and thus reduce the overall conversion of SO₂ to n.s.s. SO₄²⁻ in sea-salt particles. SO₂ mass transfer and n.s.s. SO₄²⁻ production in sea-salt particles thus become size dependent. The SO₂ mass-transfer rate, dC(D)/dt, as a function of particle diameter D, may be determined using well established formulations for gas-phase and interfacial mass transfer¹³

$$\frac{dC(D)}{dt} = \frac{P_{\text{SO}_2}}{RT} \left[\frac{12D_g}{D^2 \left(1 + \left[\lambda + \frac{4(1-\alpha)}{3\alpha} \right] \frac{2l}{D} \right)} \right] + \frac{3v\alpha}{2D} \quad (1)$$

where C(D) is the aqueous-phase concentration at the surface of the sea-salt aerosol particles as a function of D; P_{SO₂} is the gas-phase pressure of SO₂; R is the universal gas constant; T is the absolute temperature; D_g is the SO₂ gas diffusion coefficient; α is the sticking coefficient; v is the mean molecular speed of SO₂ in the gas phase; l is the mean free path length of SO₂ molecules; and λ is determined from the theory of Sahni¹⁴ for gas diffusion to a grey sphere.

MBL SO₂ concentrations¹⁵ during CASE were 2.1–2.5 nmol m⁻³ (50–60 parts per 10¹² by volume, p.p.t.v.) with [O₃] = 15–20 p.p.b.v. (ref. 16). These concentrations, their mass transfer rates, and the water volume of the sea-salt aerosol particles (yet to be determined) allow us to estimate the overall conversion rate of SO₂ to n.s.s. SO₄²⁻. The volume of water associated with sea-salt particles (the SSAW volume) may be estimated using an empirical model¹⁷ for sea-salt particle (water) growth with relative humidity. This growth curve is the mean of data for sea-spray particle growth and Atlantic maritime growth¹⁸. For the relative humidity during the Bermuda-CASE experiment (average at 150 m, 76%, and at 15 m, 91%) the resultant SSAW mass is presented in Table 1 for 10 particle size fractions.

With the limitations determined from equation (1), SO₂ mass transfer is a factor of 7–10 lower at aircraft level (150 m above

sea level, a.s.l.) and a factor 4–5 lower at ship level (15 m a.s.l.) then it would be without limitation; its magnitude is shown in the last two columns of Table 1. Note that SO_2 is preferentially transferred into the 1.5–9.5 μm diameter range on the basis of equation (1) and not on the basis of particle volume or surface area⁹. The measured n.s.s. SO_4^{2-} size distribution at 15 m a.s.l.⁷ supports the idea that O_3 -oxidized conversion of SO_2 produces much of the n.s.s. SO_4^{2-} in sea-salt particles. Given that O_3 mass-transfer limitation becomes significant (O_3 aqueous-phase transfer reduced by >10%) for particles more than 7 μm in diameter^{12,19} and that SO_2 mass-transfer peaks in the 3.5–5.5- μm diameter range (Table 1), it may be expected that n.s.s. SO_4^{2-} will be preferentially found in <6 μm sea-salt particles. Indeed, of the n.s.s. SO_4^{2-} found in sea-salt particles during Bermuda-CASE,⁷ about 65% was found in the 1–6 μm fraction.

The roughly threefold variability in SSAW mass across the 15–150 m a.s.l. height range, from 17 to 46 $\mu\text{g H}_2\text{O m}^{-3}$ (columns 2 and 3 of Table 1), is considered to encompass the uncertainty in Fitzgerald's empirical model of sea-salt particle growth with relative humidity¹². For a first calculation of SO_2 conversion in SSAW, a 15–50 $\mu\text{g H}_2\text{O m}^{-3}$ range will suffice; future studies of this mechanism should determine the percentage of H_2O within sea-salt particles as a function of particle size.

Combining SO_2 and O_3 mass transfer with aqueous O_3 oxidation in the estimated SSAW mass, we arrive at the result that 2–5% of the relatively constant 2.1–2.5 $\text{nmol SO}_2 \text{ m}^{-3}$ (during Bermuda-CASE) was converted to n.s.s. SO_4^{2-} in the SSAW every hour. Residence times for sea-salt aerosol particles are usually stated as 30 hr or more²⁰. Conservatively, a 20–30-hr residence time implies the production of 1–4 $\text{nmol n.s.s. SO}_4^{2-} \text{ m}^{-3}$. Although this range has an underlying factor-of-two uncertainty, the conversion of SO_2 by O_3 oxidation in SSAW can readily explain the 1.8–2.0 $\text{nmol n.s.s. SO}_4^{2-} \text{ m}^{-3}$ observed⁷ in sea-salt particles during Bermuda-CASE. The six-stage impactor data (A. Pszenny, personal communication) showed 72% of the sea-salt particle n.s.s. SO_4^{2-} to be in a 1–5- μm size range; Prospero and Savoie⁵ found ~80% in this size range at Barbados. Both of these data sets indicate that the predicted mass median diameter for sea-salt particle n.s.s. SO_4^{2-} is smaller than either volume or area median diameters. It is feasible that this conversion mechanism may produce as much n.s.s. SO_4^{2-} in the MBL as homogeneous SO_2 conversion, which was estimated¹⁰ to produce 4–5 nmol m^{-3} of the 9 $\text{nmol n.s.s. SO}_4^{2-} \text{ m}^{-3}$ observed in fine (accumulation mode) particles during Bermuda-CASE.

Modelling analysis of the species formed in the SSAW during

TABLE 1 Sea-salt aerosol water mass and SO_2 mass transfer

Diameter range (μm)	SSAW mass* ($\mu\text{g m}^{-3}$)		SO_2 mass transfer† ($\text{ng m}^{-3} \text{ h}^{-1}$)	
	150 m a.s.l.	15 m a.s.l.	150 m a.s.l.	15 m a.s.l.
<1.5	<0.1	<0.1	1.0	0.8
1.5–3.5	2.0	1.6	5.7	3.9
3.5–5.5	5.2	7.7	7.2	8.9
5.5–7.5	4.4	11.3	3.7	7.6
7.5–9.5	2.9	9.7	1.7	4.5
9.5–11.5	1.6	6.6	0.7	2.3
11.5–15.5	0.9	6.3	0.3	1.5
15.5–19.5	0.2	2.0	0.02	0.3
19.5–23.5	0.1	0.6	0.01	0.06
>23.5	<0.1	<0.2	<0.01	<0.02
Total	17	46	~20	~30

Results are shown as a function of particle size in the Bermuda area during 26–28 July, 1988 CASE sampling (from ref. 12).

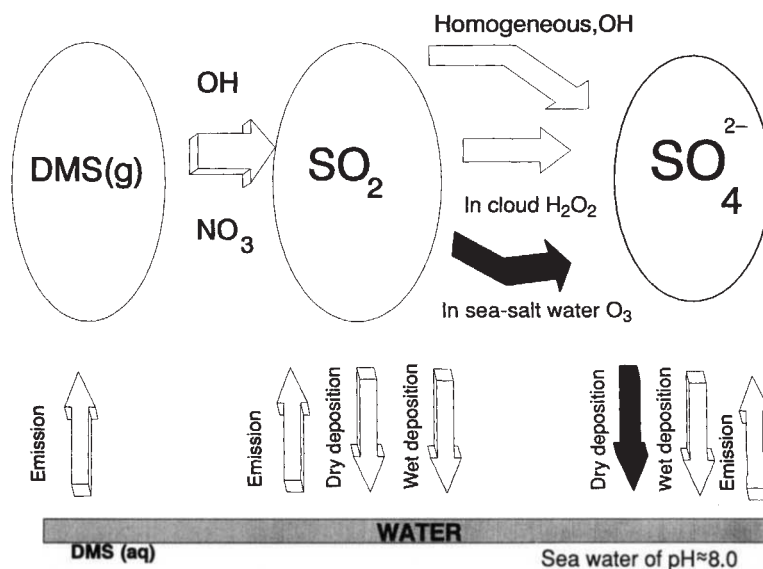
* The ambient, log-normally fitted volume distributions of the sea-salt mode were determined using the forward scattering spectrometer optical probe^{8,12}. The SSAW mass ($\rho_{\text{H}_2\text{O}}=1.0$) in 10 size fractions was then estimated using Fitzgerald's empirical model.¹⁷ Variability in hourly SSAW mass estimations during the July 26–28 period was <15%; thus, averages of 33 hourly values for each size fraction are presented.

† Using equation (1) along with the SSAW mass estimates.

Bermuda-CASE (S.P. and H.S., unpublished results) suggests that the SSAW pH should be kept at about 7, despite the production of n.s.s. SO_4^{2-} . (Only if the pH is kept above 6–6.5 will O_3 oxidation of SO_2 be rapid.) These calculations do not fully consider the ionic character of sea-derived water; Chameides and Stelson²¹ have done so, however, and found no significant reduction in SO_2 oxidation by O_3 . Roughly 60% of the n.s.s. SO_4^{2-} should be neutralized by CaCO_3 in the sea-salt particles as CO_2 is released. Another 10–30% may be neutralized by reaction with NaCl and release of HCl . Only a small fraction, if any, will be present as either $(\text{NH}_4)_2\text{SO}_4$ or NH_4HSO_4 . Indeed, during Bermuda-CASE²² and also over the equatorial Pacific Ocean²³ substantial amounts of CaSO_4 and mixed-cation sulphates have been observed to be associated with sea-salt particles.

This MBL SO_2 conversion mechanism may have important implications for a proposed linkage between the global sulphur

FIG. 1 Sulphur cycling in the marine boundary layer with emphasis on the O_3 -oxidized heterogeneous SO_2 conversion and sulphur removal pathway (DMS to MSA conversion is not shown).



and carbon cycles. Charlson *et al.*⁴ have hypothesized that enhanced DMS emissions from warmer ocean surface waters may produce new cloud condensation nuclei as the DMS converts to n.s.s. SO_4^{2-} in fine, accumulation-mode particles. The resulting sulphate haze would produce a direct radiative cooling influence and the enhanced cloud cover would produce indirect cooling. The DMS that converts to SO_2 (as opposed to methanesulphonate, MSA) may, however, convert further to n.s.s. SO_4^{2-} produced in SSAW, which, being predominantly associated with $>2\text{ }\mu\text{m}$ diameter particles, will dry-deposit at a rapid rate^{24,25}. From the observed n.s.s. SO_4^{2-} distributions during Bermuda-area CASE⁷ or Barbados⁸ sampling, we estimate that $>90\%$ of the n.s.s. SO_4^{2-} deposition comes from $>1.2\text{ }\mu\text{m}$ diameter particles. A 20–30 hr residence time, given a MBL height of 1–2 km, implies a dry deposition velocity $>1\text{ cm s}^{-1}$ (roughly twice that for SO_2), similar to the n.s.s. SO_4^{2-} dry deposition velocity needed for mass balance of sulphur species in data sets obtained in the Atlantic^{10,26} MBL.

Heterogeneous SO_2 conversion in sea-salt particles, combined with the large dry deposition rates of these particles, defines a rapid recycling pathway for sulphur from the oceans (as DMS) and back to the oceans (as n.s.s. SO_4^{2-} in SSAW). This SO_2 removal pathway is highlighted in Fig. 1. Estimated DMS²⁶ and sea-salt particle⁷ emission rates for the summertime western North Atlantic Ocean near Bermuda during CASE were such that at least half of the DMS emitted may have been converted to n.s.s. SO_4^{2-} in sea-salt particles.

Worldwide, the oceanic DMS emission rate has been estimated to be $<10^{12}\text{ mol S yr}^{-1}$ (see for example ref. 27) whereas the sea-salt particle emission rate is estimated to be 10^{15} – 10^{16} g yr^{-1} (see for example ref. 28). If all the carbonate in sea-salt particles is released in the process of producing n.s.s. SO_4^{2-} by O_3 -oxidation of SO_2 in SSAW, then $>10^{11}\text{ mol S yr}^{-1}$ (between 10% and nearly 100%) of global DMS emissions that convert to SO_2 (MSA excluded) may be rapidly returned to surface ocean waters by the pathway highlighted in Fig. 1. (Additional conversion in the SSAW, for example through metal catalysis²⁹, will increase the amount of sulphur rapidly recycled through this pathway.) Instead of enhanced haze and cloud albedo resulting from increased DMS emissions, it may be that

sulphur is more rapidly exchanged between the surface waters of the global oceans and the MBL. Heterogeneous conversion of sulphur in SSAW and subsequent dry removal from the MBL might be viewed as a 'self-cleansing' process which helps to maintain a steady state (homeostasis) for the global sulphur cycle. □

Received 22 November 1991; accepted 13 November 1992.

- Seinfeld, J. H. *Atmospheric Chemistry and Physics of Air Pollution* (Wiley, New York 1986).
- Winkler, P. in *Chemistry of Multiphase Atmospheric Systems* (ed. Jaeschke, W.) 269–298 (Springer, Heidelberg, 1986).
- Sievering, H. *et al. Atmos. Envir.* **A25**, 1479–1487 (1991).
- Charlson, R. J., Lovelock, J. E., Andreae, M. O. & Warren, S. G. *Nature* **326**, 655–661 (1987).
- Gravenhorst, G. *Atmos. Envir.* **12**, 707–713 (1978).
- Prospero, J. M. & Savoie, D. L. *AEROCE Scientific Results from Phase I, 1988–90* (Rosenstiel School of Marine & Atmos. Sci., Univ. of Miami, 1991).
- Sievering, H., Ennis, G. & Gorman, E. *Global biogeochem. Cycles* **4**, 395–405 (1990).
- Pszenny, A. A. P., Artz, R. S., Boatman, J. F. & Galloway, J. N. *Global biogeochem. Cycles* **4**, 121–132 (1990).
- Kim, Y., Sievering, H. & Boatman, J. F. *Global biogeochem. Cycles* **4**, 165–178 (1990).
- Luria, M. & Sievering, H. *Atmos. Envir.* **A25**, 1489–1496 (1991).
- Hoppel, W. A., Fitzgerald, J. W., Frick, G. M., Larson, R. E. & Mack, E. J. *J. geophys. Res.* **95**, 3659–3686 (1990).
- Sievering, H. *et al. in Proc. 5th IPSASEP Conf.* Richland, July, 1991 (eds Schwartz, S. & Slinn, W. G. N.) (in the press).
- Schwartz, S. in *Chemistry of Multiphase Atmospheric Systems* (ed. Jaeschke, W.) 415–471 (Springer, Heidelberg, 1986).
- Sahni, D. J. *nucl. Energy* **20**, 915–920 (1966).
- Luria, M. *et al. Global biogeochem. Cycles* **4**, 381–394 (1990).
- Ray, J. D., Van Valin, C. C., Luria, M. & Boatman, J. F. *Global biogeochem. Cycles* **4**, 201–214 (1990).
- Fitzgerald, J. W. *Appl. Optics* **28**, 3534–3538 (1989).
- Hänel, F. *Adv. Geophys.* **19**, 73–188 (1976).
- Gorman, E. *Mass Transfer Limitation of SO_2 and Ozone in Coarse-Mode Aerosol Particles* (Center for Environ. Sciences, Univ. of Colorado, 1992).
- Slinn, W. G. N. in *Precip. Scav., Dry Deposition, and Resusp., Vol. 2* (eds Pruppacher, H. R., Semonin, R. G. & Slinn, W. G. N.) 1361–1416 (Elsevier, New York 1983).
- Chameides, W. L. & Stetson, A. W. *J. geophys. Res.* (in the press).
- Kopcewicz, B. *et al. Atmos. Res.* **26**, 245–271 (1991).
- Andreae, M. O. *et al. Science* **232**, 1620–1623 (1986).
- Sievering, H. *J. geophys. Res.* **89**, 9679–9681 (1984).
- Slinn, S. A. & Slinn, W. G. N. *Atmos. Envir.* **14**, 1013–1016 (1980).
- Luria, M. *et al. Atmos. Envir.* **23**, 139–147 (1989).
- Leck, G. & Rodhe, H. *J. atmos. Chem.* **12**, 63–86 (1991).
- Blanchard, D. & Woodcock, A. H. *N.Y. Acad. Sci.* **338**, 330–347 (1980).
- Zhuang, G., Yi, Z., Duce, R. A. & Brown, P. *Nature* **355**, 537–539 (1992).

ACKNOWLEDGEMENTS. We thank R. Artz, W. Keene, C. Nagamoto, F. Parungo and A. Pszenny for field help and discussions during data interpretation. We also thank the crew of the NOAA ship *Mt Mitchell* as well as the NOAA Aerosol Research Section, especially L. Gunter and D. Wellman, for technical and logistical support. This work was funded by the NSF, Office of Naval Research and NOAA/ERL/ARL.

Deriving global climate sensitivity from palaeoclimate reconstructions

Martin I. Hoffert* & Curt Covey†

* Earth Systems Group, Department of Applied Science, New York University, New York 10003, USA

† Lawrence Livermore National Laboratory, Livermore, California 94551, USA

To assess the future impact of anthropogenic greenhouse gases on global climate, we need a reliable estimate of the sensitivity of the Earth's climate to changes in radiative forcing. Climate sensitivity is conventionally defined as the equilibrium surface temperature increase for carbon dioxide doubling, $\Delta T_{2\times}$. Uncertainties in cloud processes spread general circulation model (GCM) estimates of this parameter over the range $1.5 < \Delta T_{2\times} < 4.5\text{ }^\circ\text{C}$ (refs 1, 2). An alternative to model-based estimates is in principle available from the reconstruction of past climates^{3–6}, which implicitly includes cloud feedback. Here we retrieve the sensitivity of two palaeoclimates, one colder and one warmer than present, by independently reconstructing both the equilibrium surface temperature change and the radiative forcing. Our results yield $\Delta T_{2\times} = 2.3 \pm 0.9\text{ }^\circ\text{C}$. This range is comparable with estimates from GCMs and inferences from recent temperature observations and ocean

models^{7,8}. Future application of the method to additional climates in the geological record might constrain climate sensitivity enough to narrow the model uncertainties of global warming predictions.

A change in the global mean surface temperature $\Delta \bar{T}$ is normally produced by a change in net radiative forcing⁹,

$$\Delta \bar{Q} = \Delta Q_{\text{Sun}} + \Delta Q_{\text{albedo}} + \Delta Q_{\text{green}} + \Delta Q_{\text{aerosol}} \quad (1)$$

where $\Delta Q_{\text{Sun}} = (S_0/4)(1 - \alpha)(\Delta S/S_0)$ is the forcing from solar irradiance changes ΔS , $\Delta Q_{\text{albedo}} = -(S_0/4)\Delta\alpha$ the forcing from surface albedo changes $\Delta\alpha$, ΔQ_{green} (ΔCO_2 , ΔCH_4 , ...) the forcing from greenhouse gas concentration changes and $\Delta Q_{\text{aerosol}}$ (ΔSO_4 , ...) the forcing from changes in atmospheric aerosols. Present-day solar constant and planetary albedo values are $S_0 \approx 1,370\text{ W m}^{-2}$ and $\alpha \approx 0.30$. A greenhouse gas reference forcing commonly used in global warming analyses is that from atmospheric CO_2 doubling⁹, $\Delta Q_{2\times} \approx 4.4\text{ W m}^{-2}$.

The simplest energy balance for an applied radiative perturbation is $\Delta \bar{Q} = \lambda_b \Delta \bar{T}$, where $\lambda_b \approx 3.8\text{ W m}^{-2}\text{ K}^{-1}$ is the radiative damping coefficient for black-body cooling¹⁰. In that case, $\Delta T_{2\times} = \Delta Q_{2\times}/\lambda_b \approx 1.2\text{ }^\circ\text{C}$.

If feedbacks over cloud-covered parts of the Earth were neutral, an imposed radiative forcing would still induce a feedback flux per unit surface area of Earth, ΔQ_{clear} , by water vapour feedbacks in clear-sky regions. The energy balance including clear-sky feedbacks, $\Delta \bar{Q} + \Delta Q_{\text{clear}} = \lambda_b \Delta \bar{T}$, yields the clear-sky radiative damping, $\lambda_{\text{clear}} \equiv \Delta \bar{Q}/\Delta \bar{T} = \lambda_b - \Delta Q_{\text{clear}}/\Delta \bar{T}$. Radiative-convective models¹¹ supported by satellite observations¹² indicate positive feedback ($\Delta Q_{\text{clear}}/\Delta \bar{T} > 0$) from increased

infrared opacity of the troposphere as more water evaporates into air columns over warmer surfaces. This reduces radiative damping and increases climate sensitivity relative to black-body cooling. GCMs show positive water vapour feedback, and agree to within ~10% on clear-sky damping and climate sensitivity¹; $\lambda_{\text{clear}} = 2.2 \pm 0.2 \text{ W m}^{-2} \text{ K}^{-1}$, $\Delta T_{2\times} = \Delta Q_{2\times} / \lambda_{\text{clear}} = 2.0 \pm 0.2 \text{ }^{\circ}\text{C}$.

Lindzen¹³ holds that GCMs err on this point and that water vapour feedback is in fact negative ($\Delta Q_{\text{clear}} / \Delta T < 0$, because tropical cumulus towers enhanced by global warming dry the upper troposphere enough to reduce infrared opacity world-wide). In that case, $\lambda_{\text{clear}} > \lambda_b$; perhaps⁸ $\lambda_{\text{clear}} \approx 8.8 \text{ W m}^{-2} \text{ K}^{-1}$ ($\Delta T_{2\times} \sim 0.5 \text{ }^{\circ}\text{C}$).

In GCMs, the main uncertainty in $\Delta T_{2\times}$ is associated with radiative fluxes, ΔQ_{cloud} , induced by cloud formation and cloud radiative processes when the climate changes. The energy balance $\Delta \bar{Q} + \Delta Q_{\text{clear}} + \Delta Q_{\text{cloud}} = \lambda_b \Delta \bar{T}$ gives $\lambda = \Delta \bar{Q} / \Delta \bar{T} = \lambda_b - \Delta Q_{\text{cloud}} / \Delta T - \Delta Q_{\text{clear}} / \Delta T$, or

$$\lambda = \lambda_{\text{clear}} - (\Delta Q_{\text{cloud}} / \Delta T) \quad (2)$$

Figure 1 shows that λ and $\Delta T_{2\times}$ of GCMs are uncertain by a factor of three because of uncertainties in cloud radiative feedback. Even the sign of $\Delta Q_{\text{cloud}} / \Delta T$ is uncertain.

The Intergovernmental Panel on Climate Change (IPCC) in 1990 projected global warming from a transient ocean-climate model¹⁴ for a range of GCM-derived sensitivities², $\Delta T_{2\times} = 1.5$, 2.5 and 4.5 $^{\circ}\text{C}$. Four future radiative forcing scenarios, A, B, C and D, were examined, representing progressively more constrained greenhouse gas emissions. All the scenarios produced some global warming, but unconstrained emissions (Business as Usual, A; "SA90" in IPCC updates) in combination with $\Delta T_{2\times} = 1.5 \text{ }^{\circ}\text{C}$ produced about the same warming as the most

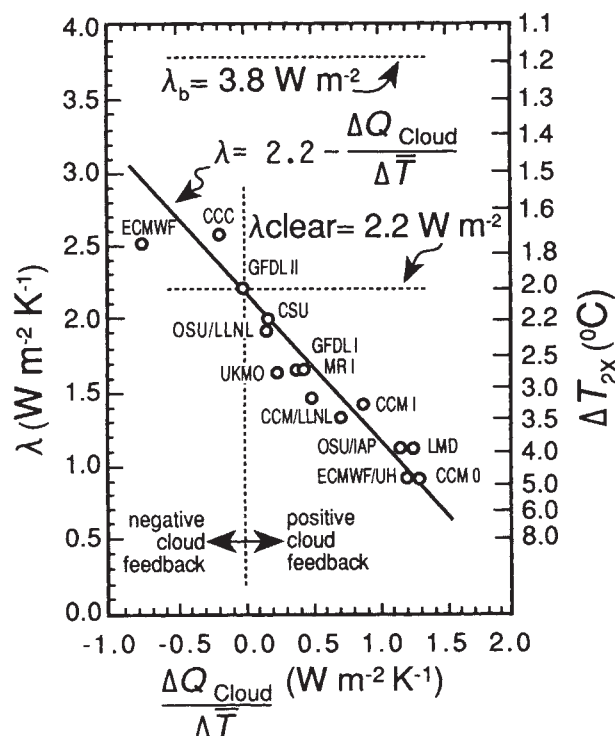


FIG. 1 Radiative damping coefficient, λ , and CO_2 doubling climate sensitivity $\Delta T_{2\times}$ plotted against the cloud radiative feedback parameter of 14 GCMs. Data points are derived from Cess *et al.*¹ where the GCM acronyms are defined (but note the different definition of the symbol λ in ref. 1). The linear correlation (solid line) implies that the differences in GCM-derived climate sensitivity arise mainly from differences in the way that cloud formation and cloud radiation physics are represented in simplified form (parameterized) in different GCMs.

TABLE 1 Sensitivity estimates of two palaeoclimates

Last Glacial Maximum (LGM) cooling ~21.5 kyr BP				
Radiative forcing (W m^{-2})				
	Component mean	Component r.m.s.	Cumulative mean	Cumulative r.m.s.
Component	$\langle \Delta Q_i \rangle$	σ_{Q_i}	$\langle \Delta Q \rangle$	$\sigma_Q = \sqrt{\sum \sigma_{Q_i}^2}$
Sun	0.0	0.2	0.0	0.2
Albedo	-3.0	0.5	-3.0	0.5
Greenhouse	-2.8	0.3	-5.8	0.6
Aerosol	-0.9	0.7	-6.7	0.9
Temperature response ($^{\circ}\text{C}$)				
			$\langle \Delta T \rangle$	σ_T
			-3.0	0.6
Climate sensitivity ($^{\circ}\text{C}$)				
			$\langle \Delta T_{2\times} \rangle$	$\sigma_{T_{2\times}}$
			2.0	0.5
Middle Cretaceous Maximum (MCM) warming ~100 Myr BP				
Radiative forcing (W m^{-2})				
	Component mean	Component r.m.s.	Cumulative mean	Cumulative r.m.s.
Component	$\langle \Delta Q_i \rangle$	σ_{Q_i}	$\langle \Delta Q \rangle$	$\sigma_Q = \sqrt{\sum \sigma_{Q_i}^2}$
Sun	-1.2	0.2	-1.2	0.2
Albedo	5.8	0.9	4.6	0.9
Greenhouse	11.1	6.7	15.7	6.8
Temperature response ($^{\circ}\text{C}$)				
			$\langle \Delta T \rangle$	σ_T
			9.0	2.0
Climate sensitivity ($^{\circ}\text{C}$)				
			$\langle \Delta T_{2\times} \rangle$	$\sigma_{T_{2\times}}$
			2.5	1.2

highly constrained emissions (D) with $\Delta T_{2\times} = 4.5 \text{ }^{\circ}\text{C}$. As D would be much more costly to implement than A, $\Delta T_{2\times}$ uncertainties can severely affect the cost projections of emission controls.

Are more accurate $\Delta T_{2\times}$ estimates possible that are independent of GCMs? One approach is to vary $\Delta T_{2\times}$ in an ocean-climate model¹⁴ to fit the historical temperature record¹⁵. The radiative forcing history from greenhouse gas buildup (~2 W m^{-2} by 1990 relative to pre-industrial) yields a relatively low sensitivity by this technique^{7,8}, $\Delta T_{2\times} \approx 1.5 \text{ }^{\circ}\text{C}$. But the calculations omit effects of increasing sulphate aerosols from fossil-fuel burning which could have produced partially compensating negative forcing (~-1 W m^{-2} by 1990 from direct scattering of sunlight and/or cloud condensation nuclei increasing marine stratus albedo¹⁶). Wigley and Raper¹⁷ found the historical temperature record could be fitted when anthropogenic sulphate aerosol forcing was included with $\Delta T_{2\times} \approx 3.4 \text{ }^{\circ}\text{C}$, whereas Schlesinger *et al.*¹⁸ estimate $\Delta T_{2\times} = 2.2 \pm 0.8 \text{ }^{\circ}\text{C}$ with aerosol forcing. Apart from uncertainties in transient forcing and ocean mixing, natural (uncertained) variability during the historical temperature period can affect climate sensitivity derived by this method. This is minimized by working with the long-term average palaeoclimates whose forcings and responses are large.

We sought to derive $\Delta T_{2\times}$ for the Last Glacial Maximum (LGM) about twenty thousand years before present (21.5 kyr BP) and the Mid-Cretaceous Maximum (MGM), about 100 Myr BP, periods of large negative and positive climatic perturbation relative to today¹⁹. Besides the fast feedbacks included in $\Delta T_{2\times}$ (timescales < 1 yr), 'slow' feedbacks (timescales of decade or more) can come into play involving greenhouse gas concentrations, ocean circulation changes⁵ and icecap-bedrock dynamics²⁰. Following equilibrium climate models^{21,22}, we treat fast feedbacks as part of the response, slow feedbacks as direct radiative forcing.

For the LGM, we estimated a mean solar radiative forcing of zero with some residual uncertainty based on 'solar-cycle' variations of Sun-like stars²³. Surface albedo forcing was estimated from changes in glacial ice, vegetation and topography^{3,22}. Greenhouse forcing was from IPCC radiative parameterizations for pre-industrial concentrations⁹, $(\text{CO}_2)_0 = 279 \text{ p.p.m.v.}$,

(CH₄)₀ = 790 p.p.b.v., and LGM concentrations from the Vostok ice core^{3,24}, CO₂ = 195 p.p.m.v., CH₄ = 350 p.p.b.v. Aerosol forcing was based on LGM sulphate particulate loading in ice cores²⁵ with numerical values from Harvey's²⁶ LGM base case reduced by a factor of 2 to 10 (ref. 27). For the MCM, a slightly negative solar forcing was estimated based on the Sun brightening by 0.5% per 100 Myr (ref. 28; $\pm 10\%$ uncertainty). Surface albedo forcing was based on calculations of the present Earth relative to the darker (ice-free and more ocean-covered) Earth 100 Myr BP with r.m.s. deviations from uncertainties in surface vegetation and continental boundaries²⁹. MCM greenhouse forcing was computed from the parameterization of Kiehl and Dickinson³⁰ for atmospheric palaeo-CO₂ concentrations 2–11 times present levels estimated from several independent reconstructions^{31–35}. Values are given in Table 1. Component uncertainties are treated as r.m.s. deviations σ_{Q_i} about the means $\langle \Delta Q_i \rangle$ of independent random variables ΔQ_i . Also in Table 1 are the cumulative forcing mean and r.m.s. uncertainty computed from³⁶ $\langle \Delta Q \rangle = \sum \langle \Delta Q_i \rangle$ and $\sigma_Q = (\sum \sigma_{Q_i}^2)^{1/2}$. Our net forcing estimates are $-6.7 \pm 0.9 \text{ W m}^{-2}$ (LGM) and $15.7 \pm 6.8 \text{ W m}^{-2}$ (MCM).

We estimated the statistical mean and r.m.s. global surface temperature changes, $\langle \Delta T \rangle$ and σ_T , from LGM and MCM palaeotemperature reconstructions as follows. Figure 2a shows distributions of zonal mean temperature change reconstructed for the Mid-Holocene ($\sim 5\text{--}6 \text{ kyr BP}$), Eemian ($\sim 125 \text{ kyr BP}$)

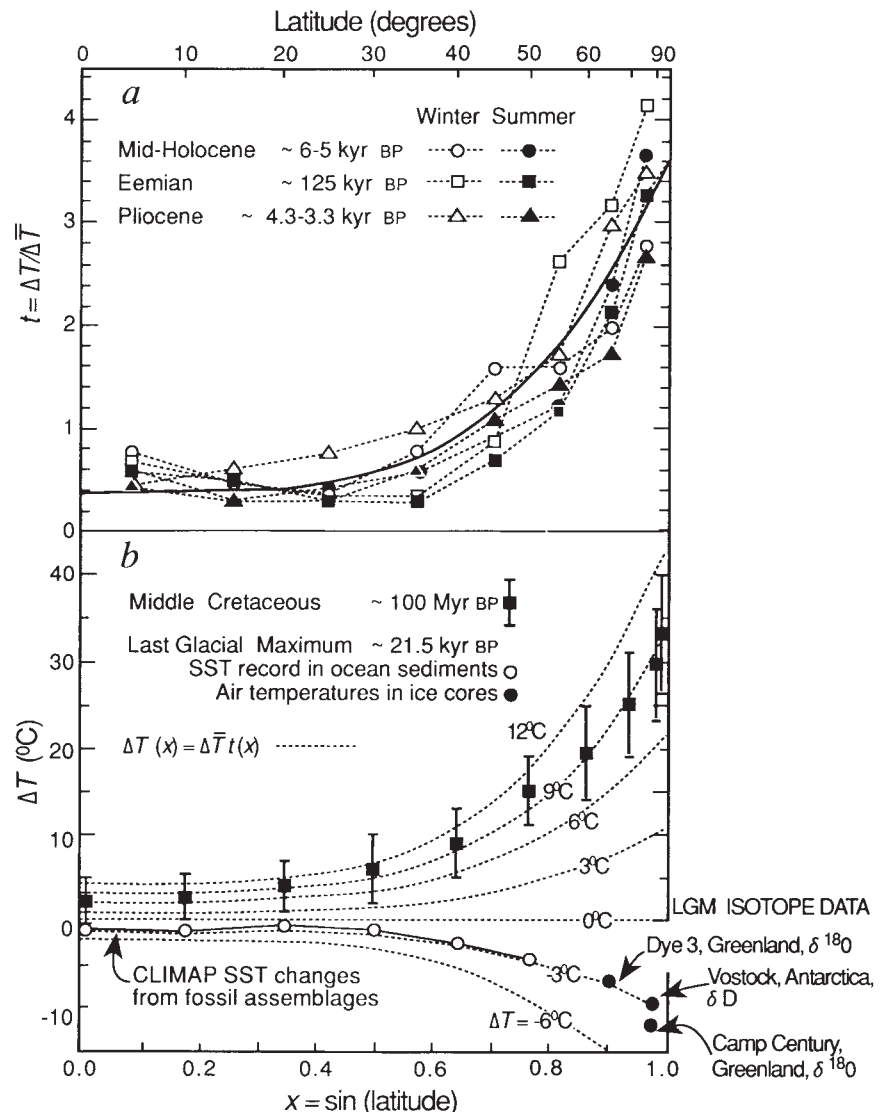
and Mid-Pliocene ($\sim 3.3\text{--}4.3 \text{ Myr BP}$) periods normalized to their global mean temperature changes⁵. Budyko and Izrael⁶ hypothesized that such normalized surface temperature changes are 'universal'. The least-squares curve fit

$$t(x) = \Delta T(x)/\Delta \bar{T} = 0.36 + 3.21x^4 \quad (3)$$

where $x = \sin(\text{latitude})$ fits the data of Fig. 2a well (correlation coefficient $r^2 = 0.89$), although it is not obvious that it describes temperature distributions of all periods³⁷. A counterexample (not shown) may have occurred in the Eocene when sea surface temperatures were apparently cooler in the tropics and warmer at high latitudes³⁸.

Figure 2b shows that Equator-to-pole temperature changes relative to present of both the LGM and MCM are described well by equation (3). The LGM temperature change distribution was synthesized from CLIMAP sea surface temperature (SST) changes equatorward of sea ice^{21,22,39} and high-latitude air temperature changes derived from ice cores. The surface air temperature change at Vostok station on the Antarctic plateau is from deuterium isotope fractionation (δD)²⁴. Surface cooling relative to present at Dye 3 and Camp Century, Greenland, was derived from oxygen isotope fractionation ($\delta^{18}O$)⁴⁰ assuming a change of $\sim 1.0\%$ $\delta^{18}O$ per degree centigrade⁴¹. Equator-to-pole temperature distributions in Fig. 2b are meant to represent long-term equilibrium climates. The $\sim 9^\circ\text{C}$ cooling predicted by

FIG. 2 a, Longitudinally averaged surface temperature changes from the present-day climate, normalized by dividing by globally averaged temperature changes. Points connected by dotted lines are based on palaeodata from the mid-Holocene, Eemian and Pliocene warm periods. The solid line is a least-squares fit to this data for a fourth-order polynomial in the sine of latitude (equation (3)). b, Longitudinally averaged surface temperature changes from the present-day climate (average of the northern and southern hemispheres). Points are derived from palaeodata for the mid-Cretaceous maximum warming and Last Glacial Maximum cooling (see text). Lines show the best-fit normalized temperature curve of a multiplied by the indicated selection of positive and negative global mean temperature changes. Comparison of the lines with a set of points gives the range of global mean temperature change that best fits the data.



the $\Delta\bar{T} = -3^\circ\text{C}$ curve at the latitude of Summit, Greenland (72°N) is a reasonable time-average of alternating mild (-7°C) and severe (-12°C) LGM glacial climates observed at this station⁴². MCM data in Fig. 2b are surface temperature reconstructions of Barron and Washington⁴³ minus the present surface air temperature distributions. The points shown are an average response of the two hemispheres, with the error bars showing upper and lower bound (not r.m.s.) uncertainties. Comparison of this data with the $\Delta T(x) = \Delta\bar{T}t(x)$ curves, also shown in Fig. 2b for different global mean temperature changes, yielded our $\Delta\bar{T}$ estimates, $-3.0 \pm 0.6^\circ\text{C}$ (LGM) and $9.0 \pm 2.0^\circ\text{C}$ (MCM).

The climate sensitivity mean and standard deviation in Table 1 were estimated from

$$\langle\Delta T_{2x}\rangle \approx \frac{\Delta Q_{2x}\langle\Delta T\rangle}{\langle\Delta Q\rangle}$$

$$\sigma_{T_{2x}} \approx \langle\Delta T_{2x}\rangle \sqrt{\left(\frac{\sigma_Q}{\langle\Delta Q\rangle}\right)^2 + \left(\frac{\sigma_T}{\langle\Delta T\rangle}\right)^2} \quad (4)$$

The approximation assumes statistically independent radiative forcing and temperature estimates and $(\sigma_Q/\langle\Delta Q\rangle)^2$, $(\sigma_T/\langle\Delta T\rangle)^2 \ll 1$. (Small errors associated with nonlinearities and uncertainties in ΔQ_{2x} are neglected.) Our ΔT_{2x} climate sensitivities are then $2.0 \pm 0.5^\circ\text{C}$ (LGM) and $2.5 \pm 1.2^\circ\text{C}$ (MCM).

Although our method retrieves a Pleistocene $\Delta\bar{T} \approx -3^\circ\text{C}$, Lorius *et al.*³ cite LGM global cooling of ~ -4 to -5°C . If the polar cooling estimated from ice cores is roughly correct, a global mean cooling of -4 to -5°C implies that the weak ($<1^\circ\text{C}$) CLIMAP sea surface cooling from the Equator to 30 degrees latitude is a considerable underestimate (see Fig. 2b). Preliminary data from palaeothermometers in groundwater⁴⁴ and isotopic and micropalaeontological records in ocean sediment cores⁴⁵ admit the possibility of LGM tropical and mid-latitude cooling $>3^\circ\text{C}$. A recalibrated, colder CLIMAP data set along with consistent land data could increase our LGM sensitivity estimate—perhaps to the mid-point of the IPCC range², $\langle\Delta T_{2x}\rangle \approx 3.0^\circ\text{C}$.

Still, it is remarkable that independently estimated sensitivities of the LGM and MCM are so close to each other and to the $\Delta T_{2x} \approx 2.2^\circ\text{C}$ that Schlesinger *et al.*¹⁸ obtained by fitting the historical temperature record to the output of an ocean-climate model forced by anthropogenic greenhouse gases and aerosols. The 'best guess' $\Delta T_{2x} \approx 3.4^\circ\text{C}$ of Wigley and Raper¹⁷, obtained with slightly different assumptions about the net radiative forcing history, is closer to the 'colder' LGM. In general, ΔT_{2x} can change as new palaeoclimate estimates are added to the statistical data base. More focused analysis of palaeodata along with palaeo-reconstructions of additional periods could narrow the uncertainties further.

The numbers are already good enough to rule out a very low sensitivity implied by Lindzen¹³. For example, $\Delta T_{2x} \approx 0.5^\circ\text{C}$ requires relatively large, and nonphysical, forcings ($\sim -26\text{ W m}^{-2}$ to chill the LGM, $\sim +80\text{ W m}^{-2}$ to heat the MCM). The climate fluctuations imprinted in the geological record could not have occurred if sensitivity was so sluggish. Interpolation of the IPCC's Business-as-Usual scenario curves² to our empirical sensitivity estimates for a variety of choices of the ocean model's adjustable parameters, and our own calculations with the Hoffert *et al.* ocean model¹⁴, indicate that global warming by the end of the next century will be 3 – 4°C . Such a warming is unprecedented in the past million years, and represents a secular climatic change much faster than previously experienced by natural ecosystems during glacial-interglacial transitions. $\square$

4. Hansen, J. E. & Lacis, A. A. *Nature* **346**, 713–719 (1990).
5. MacCracken, M. C. *et al.* (eds) *Prospects for Future Climate: A Special US/USSR Report on Climate and Climate Change* (Lewis, Chelsea, Michigan, 1990).
6. Budyko, M. I. & Izrael, Y. A. *Anthropogenic Climate Change*, 277–318 (Univ. of Arizona Press, Tucson, 1991).
7. Wigley, T. M. L. & Raper, S. C. B. in *Climate Change: Science, Impacts and Policy* (eds Jäger, J. & Fergusson, H. L.) 231–242 (Cambridge Univ. Press, 1992).
8. Schlesinger, M. E. & Jiang, X. *Nature* **350**, 219–221 (1991).
9. Shine, K. P. *et al.* in *Climate Change: The IPCC Scientific Assessment* (eds Houghton, J. T., Jenkins, G. & Ephraums, J. J.) 41–68 (Cambridge Univ. Press, 1990).
10. Mitchell, J. F. B. *Rev. Geophys.* **27**, 115–139 (1989).
11. Manabe, S. & Weatherland, R. T. *J. Atmos. Sci.* **24**, 241–259 (1967).
12. Raval, A. & Ramanathan, V. *Nature* **342**, 758–761 (1989).
13. Lindzen, R. S. *Bull. Am. Met. Soc.* **71**, 288–299 (1990).
14. Hoffert, M. I. *et al.* *J. geophys. Res.* **85** (C11) 6667–6679 (1980).
15. Wigley, T. M. L. & Raper, S. C. B. in *Climate and Sea Level Changes: Observations, Projections and Implications* (eds Warrick, R. A., Barrow, E. M. & Wigley, T. M. L.) (Cambridge Univ. Press, in the press).
16. Charlson, R. J. *et al.* *Science* **255**, 423–430 (1992).
17. Wigley, T. M. L. & Raper, S. C. B. *Nature* **357**, 293–300 (1992).
18. Schlesinger, M. E., Jiang, X. & Charlson, R. J. in *Climate Change and Energy Policy* (eds Rosen, L. & Glasser, R.) 75–108 (American Institute of Physics, New York, 1992).
19. Crowley, T. J. & North, G. *Paleoclimatology* (Oxford Univ. Press, 1991).
20. Deblonde, G. & Peltier, W. R. *Clim. Dynam.* **5**, 103–110 (1990).
21. Manabe, S. & Broccoli, A. J. *J. Atmos. Sci.* **42**, 2643–2651 (1985).
22. Harvey, L. D. D. *J. geophys. Res.* **94** (D10), 12873–12884 (1989).
23. Baliunas, S. & Jastrow, R. *Nature* **348**, 520–523 (1990).
24. Chappellaz, J. *et al.* *Nature* **345**, 127–131 (1990).
25. Delmas, R. J. & Legrand, M. in *The Environmental Record in Glaciers and Ice Sheets* (eds Oeschger, H. & Langway, C. C.) 319–341 (Wiley, New York, 1989).
26. Harvey, L. D. D. *Nature* **334**, 333–335 (1988).
27. Anderson, T. L. & Charlson, R. J. *Nature* **345**, 393 (1990).
28. Newman, M. J. & Rood, R. T. *Science* **198**, 1035–1037 (1977).
29. Thompson, S. L. & Barron, E. J. *J. Geol.* **89**, 143–167 (1981).
30. Kiehl, J. T. & Dickinson, R. E. *J. geophys. Res.* **92** (D3), 2991–2998 (1987).
31. Budyko, M. I., Ronov, A. B. & Yanshin, A. L. *History of the Earth's Atmosphere* (Springer, New York, 1987).
32. Berger, W. H. & Spitz, A. *Paleoceanography* **3**, 401–411 (1988).
33. Popp, B. N. *et al.* *Am. J. Sci.* **289**, 436–454 (1989).
34. Berner, R. A. *Science* **249**, 1382–1386 (1990).
35. Berner, R. A. *Nature* **358**, 114 (1992).
36. Minkoff, J. *Signals, Noise & Active Sensors* (Wiley, New York, 1992).
37. Crowley, T. J. *J. Clim.* **3**, 1282–1292 (1990).
38. Barron, E. J. *Paleoceanography* **2**, 729–739 (1987).
39. CLIMAP Project Members *Science* **191**, 1131–1137 (1976).
40. Dansgaard, W. *et al.* in *Climate Processes and Climate Sensitivity* (eds Hansen, J. E. & Takahashi, T.) 288–298 (American Geophysical Union, Washington DC, 1984).
41. Bradley, R. S. *Quaternary Paleoclimatology*, 132 (Unwin Hyman, Boston, 1985).
42. Johnsen, S. J. *et al.* *Nature* **359**, 311–313 (1992).
43. Barron, E. J. & Washington, W. M. in *The Carbon Cycle and Atmospheric CO₂: Natural Variations Archean to Present* (eds Sundquist, E. T. & Broecker, W. S.) 546–553 (American Geophysical Union, Washington DC, 1985).
44. Stute, M. *et al.* *Science* **256**, 1000–1003 (1992).
45. Emiliani, C. *Science* **257**, 1462 (1992).

ACKNOWLEDGEMENTS. We thank T. Wigley for suggestions, M. Budyko, K. Caldeira, A. Lapienis and D. Rind for discussions, and D. Lischner and J. M. Schantz for programming assistance. This work was supported at NYU by the U.S. Department of Energy (DOE), Atmospheric and Climate Research Division, and the National Institute for Global Environmental Change; and performed in part under the auspices of the DOE Environmental Sciences Division by the Lawrence Livermore National Laboratory.

Patterns of species diversity in the deep sea as a function of sediment particle size diversity

Ron J. Etter & J. Frederick Grassle*

Biology Department, University of Massachusetts, Boston, Massachusetts 02125, USA

* Institute of Marine and Coastal Sciences, Rutgers University, New Brunswick, New Jersey 08903, USA

UNDERSTANDING the processes that generate and maintain patterns of species diversity is a major focus of contemporary ecological and evolutionary research. In the deep sea, species diversity varies geographically and bathymetrically^{1–3}, and may attain levels that rival tropical communities⁴. Many hypotheses have been proposed concerning the forces that shape patterns of species diversity in the deep sea⁵, but so far it has not been possible to relate these patterns to potential causes in a direct quantitative way. The nature of sediments should be important in structuring deep-sea communities because deposit feeders rely on the sediments for nutrition and comprise most of the organisms in the deep sea⁶. The composition of soft sediment communities is influenced by sediment particle

TABLE 1 Correlations between species diversity and ten environmental variables (r_1) for each geographic region and all regions combined

	North			Mid			South			Combined		
	r_1	r_2	r_3	r_1	r_2	r_3	r_1	r_2	r_3	r_1	r_2	r_3
Depth	0.653***	0.031	—	0.243***	0.169*	—	0.146	0.355***	—	0.430***	0.107*	—
Sediment and water	0.174*	0.119	0.279***	0.274***	—	0.214**	0.603***	0.480***	0.656***	0.015	0.088	0.099*
Sediment	0.323***	0.163*	0.345***	0.217***	0.141*	0.158*	0.657***	—	0.702***	0.149**	0.139**	0.207***
Silt	0.665***	—	0.398***	0.233***	0.134*	0.105	0.321**	0.292*	0.295*	0.545***	—	0.424***
Clay	0.626***	0.093	0.341***	0.229***	0.126	0.158*	0.368**	0.424**	0.358**	0.465***	0.012	0.387***
Mean ϕ	0.616***	0.023	0.316***	—	—	—	0.253*	0.434***	0.209	0.521***	0.137*	0.420***
Sorting coefficient	0.338***	0.248***	0.356***	—	—	—	0.390***	0.213	0.432***	0.311***	0.260***	0.375***
% C	0.516***	0.098	0.168*	0.244***	0.075	0.123	0.425***	0.235	0.488***	0.358***	0.099*	0.305***
% H	0.375***	0.166*	0.043	0.196**	0.076	0.136*	0.643***	0.544***	0.657***	0.359***	0.117*	0.303***
% N	0.514***	0.125	0.159*	0.25***	0.087	0.137*	0.212	0.361**	0.223	0.421***	0.019	0.292***

Columns r_2 and r_3 represent correlations after controlling for sediment diversity or depth, respectively. In each analysis we held constant the measure of sediment diversity that explained the maximum variance in species diversity. The actual variable controlled for in each analysis is indicated by a dash. All relationships are linear except depth, which is parabolic. Sediment and water, Sediment, Silt, and Clay refer to the particle size diversity of those sediment fractions; Mean ϕ , and Sorting coefficient are the standard measures of mean and variance of particle size; per cent C, H, and N refer to the content of these elements in the organic fraction of the sediments. Sediment and water diversity was calculated using the per cent by weight of water in the sample as one of the entries in the Shannon–Wiener diversity index. Silt diversity was based on four ϕ fractions between 4ϕ and $<8\phi$; clay diversity was based on three ϕ fractions between 8ϕ and $>10\phi$. Mean ϕ and the sorting coefficient were not available for Mid boxcores. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

size^{7,8}. Shallow-water deposit feeders selectively ingest particular size fractions of the sediments^{9,10} and there are interspecific differences in particle size preference^{11–13}. Partitioning of sediments with respect to size may be more likely in the deep sea if there is strong selection for macrophagy as a result of reduced food supply and digestive constraints imposed by feeding on deposits¹⁴; macrophagy would permit species to ingest selectively the more labile components of the sediments. If deposit feeders in the deep sea partition the sediments with respect to size, species diversity may in part be a function of sediment particle size diversity. Also, sediment particle size diversity may reflect habitat complexity because the organisms live on or within the sediments^{15–21}. Here we show that species diversity is a significant positive function of sediment particle size diversity. The relationship seems to be scale-invariant, accounting for a similar proportion of the variance at inter-regional, regional and local scales. Bathymetric patterns of species diversity also appear to be largely attributable to changes in sediment characteristics with depth. These results suggest that sediment diversity has an important role in determining the number of species within a community and identify a direct environmental factor that potentially influences species diversity in the deep sea.

To test for a relationship between species diversity and sediment particle size diversity (referred to as sediment diversity) we collected quantitative benthic samples from bathyal depths (250–3,029m) in the western North Atlantic. We used 0.25m² boxcores from the Atlantic Continental Slope and Rise study^{22,23} because the sediment characteristics and the faunal diversities were quantified from each boxcore and because these data represent the most extensive quantitative samples of deep-sea community structure. The samples were collected from three bathyal regions along the east coast of the United States. The north Atlantic, southeast of Massachusetts; the mid-Atlantic, off the coasts of New Jersey and Delaware; and the south Atlantic, east of North and South Carolina. These regions will hereafter be referred to as the North, Mid, and South, respectively. A total of 558 boxcores were analysed and contained 272,009 individuals distributed among 1,597 species.

Details of the methods used in collecting and analysing the boxcores have been described^{22,23}. Fauna were sieved from the central nine subcores (0.09 m²) of each boxcore using a 0.3-mm sieve. Macrofaunal diversity for each boxcore was calculated using Hurlbert's²⁴ expected number of species, normalized to 100 individuals, and incorporates all taxa. For each boxcore, particle size diversity of disaggregated sediments was calculated from the per cent dry weight of each ϕ size class (12 size classes between -1 and $>10\phi$) using the Shannon–Wiener diversity

index. Because deposit feeders generally consume the smaller (silt) size fractions^{10,25}, sediment diversity was calculated for the silt fraction, clay fraction, and for all fractions with and without water. The sorting coefficient, the standard measure for variance in grain size, was also used as an environmental variable.

We explored the relationship between species diversity and ten environmental variables using standard linear and nonlinear least-squares methods. Species diversity was highly correlated with most of the environmental variables (r_1 in Table 1). For each region and for all data combined, species diversity is most strongly correlated with sediment diversity (or one of the sediment fractions), which generally accounted for more than 40% of the variance. Figure 1a and b depicts the relationship between species diversity and either depth or silt diversity for the North. Species diversity is a nonlinear function of depth with a peak in diversity at intermediate depths and is a linear function of silt diversity. The parabolic relationship between species diversity and depth is typical of previous studies of bathymetric variation in deep-sea community structure^{2,3}.

We used partial regression techniques to better characterize the relationship between species diversity and the two most important environmental variables, silt diversity and depth. If silt diversity is controlled statistically, the depth-related patterns in species diversity no longer exist (Fig. 1c). In contrast, if depth is held constant, there remains a significant relationship between the residuals, although it becomes nonlinear (Fig. 1d). This asymmetry suggests that the depth-related patterns of species diversity reflect the effects of silt diversity and are a consequence of changes in sediment characteristics with depth.

The Mid, South and combined data show similar patterns but the details are more complicated (Table 1). For the Mid and combined data, both depth and species diversity remain significant when the alternate variable is controlled, but the significance levels are asymmetric. If sediment diversity is held constant (r_2), depth is only marginally significant; if depth is held constant (r_3), sediment diversity remains highly significant. In the south, species diversity is initially uncorrelated with depth. But if sediment diversity is statistically controlled, depth becomes important in accounting for the variance in the residuals. When depth is held constant, the relationship between species diversity and sediment diversity actually improves. Similar analyses using species richness with variation in density factored out as the dependent variable gave similar patterns.

These results demonstrate that much of the variability in species diversity, geographically and bathymetrically, seems to

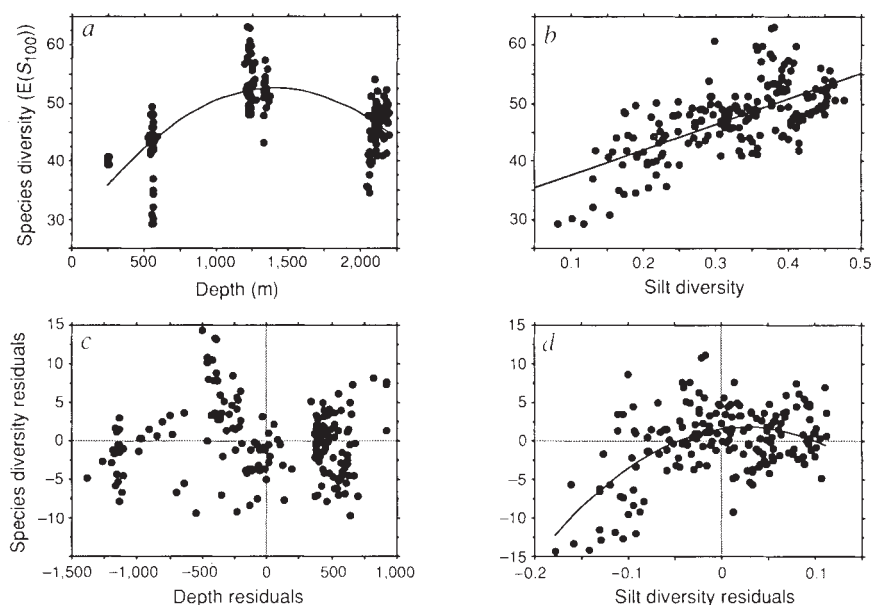


FIG. 1 The relationship between species diversity and *a*, depth or *b*, silt diversity for the North samples. Species diversity was normalized to 100 individuals. Silt diversity refers to the particle size diversity of the silt fraction of the sediments. The relationship between the residuals after the effects of silt diversity (*c*) or depth (*d*) are statistically removed. Regression

lines were fitted to the data using typical least-squares procedures and are all highly significant ($P \leq 0.0001$). No line is shown in (*c*) because the relationship was not significant ($P = 0.67$). $E(S_{100})$ is the expected number of species in each sample when normalized to 100 individuals²⁴.

be related to changes in sediment characteristics. Moreover, this explanation of species diversity appears to be just as effective at inter-regional (combined analysis), regional (within North and South) and local (Mid samples with little depth variability) scales. Although the correlation is highly significant, sediment diversity is least successful in accounting for the variance in species diversity at the Mid site. This can probably be attributed to the fact that the Mid samples were also the most homogeneous with respect to sediment diversity and species diversity.

Species diversity in intertidal soft-sediment communities also seems to be strongly correlated with variability in particle and food diversity²⁵. The similarity in the relationship between species diversity and sediment diversity in shallow- and deep-water communities suggests that this phenomenon may be a general feature of soft-sediment communities.

Regardless of the strength of a relationship or how universal it seems, correlations do not imply causality. Indeed, one could argue that highly diverse communities create more diverse sediment characteristics. For example, many of the particles have a biological origin, burrowing organisms may redistribute sediments of particular size classes, and biogenic structures create more variable hydrodynamic regimes that may induce the depo-

sition of a greater range of particle sizes^{26–28}. It is also possible that the correlation reflects more important proximal factors such as nutrient availability, hydrodynamics, microtopography or bioturbation.

The exact nature of the relationship between sediment diversity and species diversity and the mechanisms that underlie it can only be determined through direct experimentation. Experimental testing of this putative relationship should be straightforward, and in this sense alone offers advantages over previous plausible but untestable hypotheses. These results are sufficiently compelling to warrant a more detailed investigation of the role of sediment diversity in controlling the spatiotemporal variability in deep-sea species diversity.

The deep sea covers nearly two-thirds of the Earth's surface, yet the processes that generate and maintain species diversity in this vast and complex environment are poorly understood. Our work now identifies a direct environmental factor that is potentially responsible for generating spatiotemporal patterns of species diversity in the deep sea. Because similar findings have been documented in shallow water, the effects of sediment diversity on species diversity may be a ubiquitous feature of soft-sediment communities. □

Received 16 June; accepted 13 October 1992.

- Sanders, H. L. *Am. Nat.* **102**, 243–282 (1968).
- Rex, M. A. *Rev. Ecol. Syst.* **12**, 331 (1981).
- Rex, M. A. in *The Sea* (ed. Rowe, G. T.) 453–472 (Wiley, New York, 1983).
- Grassle, J. F. & Maciolek, N. J. *Am. Nat.* **139**, 313–341 (1992).
- Gage, J. D. & Tyler, P. A. *Deep Sea Biol.* 1–522 (Cambridge Univ. Press, UK, 1991).
- Sanders, H. & Hessler, R. *Science* **163**, 1419–1423 (1969).
- Gray, J. S. *Oceanogr. mar. Biol. A. Rev.* **12**, 223–261 (1974).
- Gray, J. S. *The Ecology of Marine Sediments* (Cambridge Univ. Press, UK, 1981).
- Taghon, G. L. *Oecologia* **52**, 295–304 (1982).
- Self, R. F. L. & Jumars, P. A. *J. mar. Res.* **46**, 119–143 (1988).
- Fenchel, T., Kofoed, L. H. & Lappalainen, A. *Mar. Biol.* **30**, 119–128 (1975).
- Fenchel, T. & Kofoed, L. H. *Oikos* **27**, 19–32 (1976).
- Whitlatch, R. B. *J. mar. Res.* **38**, 743–765 (1980).
- Jumars, P. A., Mayer, L. L., Deming, J. W., Baross, J. A. & Wheatcroft, R. A. *Phil. Trans. R. Soc. Lond. A* **331**, 85–101 (1990).
- MacArthur, R. & MacArthur, J. *Ecology* **42**, 594–598 (1961).
- Murdoch, W. W., Evans, F. C. & Peterson, C. H. *Ecology* **53**, 819–829 (1972).
- Pianka, E. R. *Evolutionary Ecology* 1–397 (Harper & Row, New York, 1978).

- Pielou, E. C. *Ecological Diversity* (Wiley, New York, 1975).
- Pianka, E. R. *Ecology* **48**, 333–351 (1967).
- Heck, K. L. & Wetstone, G. S. *J. Biogeogr.* **4**, 135–142 (1977).
- Dean, R. L. & Connell, J. H. *J. exp. mar. Biol. Ecol.* **109**, 195–215 (1987).
- Blake, J. A. et al. *Study of Biological Processes on the US South Atlantic Slope and Rise Phase 2*. (US Dept of Interior Report, Minerals Management Service, Washington DC, 1987).
- Maciolek, N. J. et al. *Study of Biological Processes on the US Mid-Atlantic Slope and Rise, Phase 2*. (US Dept of Interior Report, Minerals Management Service, Washington DC, 1987).
- Hurlbert, S. H. *Ecology* **52**, 477–587 (1971).
- Whitlatch, R. B. *J. exp. mar. Biol. Ecol.* **53**, 31–45 (1981).
- Johnson, R. J. *mar. Res.* **32**, 313–330 (1974).
- Rhoads, D. C. & Boyer, L. F. in *Animal–Sediment Relations: The Biogenic Alteration of Sediments* (eds McCall, P. L. & Tevesz, M. J. S.) 3–52 (Plenum, New York, 1982).
- Nowell, A. R. M., Jumars, P. A. & Fauchald, K. *Limnol. oceanogr.* **29**, 645–649 (1984).

ACKNOWLEDGEMENTS. We thank all the members involved in the Atlantic Continental Slope and Rise project: the Minerals Management Service of the US Department of Interior for supporting the collection and analysis of the boxcores; J. Levinton, R. Hessler, R. Petrecca, M. Rex, G. Taghon, D. Thistle, R. Whitlatch and S. Woodin for discussion; and J. Levinton, M. Rex, D. Thistle and R. Whitlatch for comments on the manuscript.

Dispersal of juveniles and variable recruitment in sessile marine species

Steven D. Gaines & Mark D. Bertness

Program in Ecology and Evolutionary Biology, Box G-W, Brown University, Providence, Rhode Island 02912, USA

MARINE species commonly have broadly dispersing juveniles called larvae. Their return to the adult populations is highly variable¹⁻³, often generating large fluctuations in population size⁴⁻⁶, yet the causes of the variation are poorly understood. Historically, attention has been focused on the roles of variable reproductive output by adults and variable mortality during larval development^{7,8}. The limited success of these factors as general explanations prompted a more recent focus on the influence of variable transport of the larvae⁹⁻¹³. Here we show that nearly a decade of settlement variation of the barnacle, *Semibalanus balanoides* (L.), closely matched predictions based solely on a transport hypothesis: differences in transport generate recruitment variation by determining whether larvae complete development near a favourable habitat. The irregular nature of coastlines, particularly the presence of bays and estuaries, generates substantial regional variation in coastal transport that may generate correspondingly large variation in recruitment to marine populations.

The prediction that failure of larvae to return to shore is a common cause of variable recruitment in coastal species follows

from numerous observations (such as location of fish spawning grounds¹⁰, correlations between recruitment and shifts in wind direction^{9,10,12,14,15}, behaviour of larvae^{16,17}). If the prediction is correct, then settlement ought to vary consistently with variation in transport. With this hypothesis in mind, we focused on a common feature of coastlines that predictably alters patterns of retention, bays. Rates of exchange between bays and nearby coastal waters vary because of differences in tidal amplitude, bay morphology, river input and winds¹⁸. The magnitude of variation among bays and among years is extensive and should provide strong tests of the significance of larval loss due to offshore transport.

Retention of larvae in bays has been studied for decades, but largely from the perspective of strictly estuarine species that must remain in or return to estuaries to survive¹⁹⁻²¹. The issue is changed substantially if we consider more cosmopolitan species that occur in bays and estuaries but are not restricted to them. For such species, failure to return to estuarine habitats is no longer synonymous with death. Rather, export merely subjects larvae to the same physical conditions and transport processes as larvae released from coastal habitats. If transport of larvae away from the coastline is a major cause of settlement variation, then the potential retentive characteristics of bays should establish gradients of recruitment for bays versus open coast habitats. Similarly, recruitment rates should track the large variation in retention among years and bays.

To test these predictions we studied the intertidal barnacle, *Semibalanus balanoides*, at coastal and embayed sites in Rhode Island, USA. This species is ideally suited to addressing these problems because: (1) it is widely distributed both in bays and

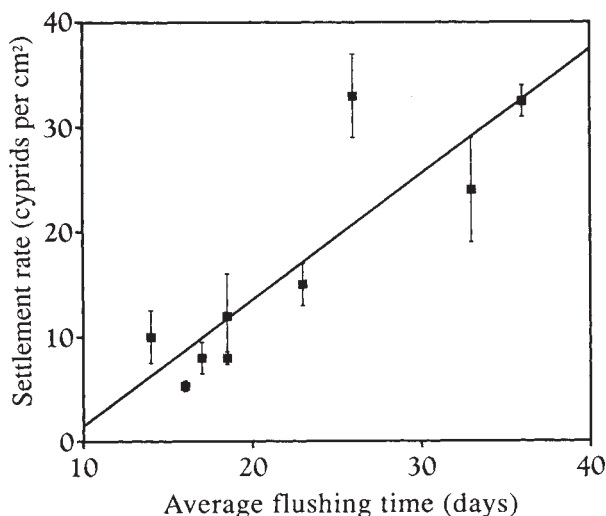


FIG. 1 Correlation between annual shoreline settlement of the barnacle *Semibalanus balanoides* and the estimated flushing time for Narragansett Bay, Rhode Island. Settlement rates were measured from photographs of 100-cm² quadrats ($n=16$ per year) taken at the end of the settlement season (that is they measure the total accumulation of settlers for a given year). Flushing times are estimates for the period January 15 to March 15, the normal interval of larval presence in Narragansett Bay. Error bars are 95% confidence limits. Flushing times are a measure of exchange rates between bay and coastal waters. They can be measured either by releasing a known quantity of material (such as dyes) into the bay and measuring the decline in concentration with time or by following the dynamics of natural constituents of the bay (such as fresh water). The volume of fresh water in a bay can be estimated from its salinity structure, and the source of fresh water is primarily river flow. Assuming that the fresh water content is in a steady state, the flushing time can be expressed simply as the time required for the river flow to equal the volume of fresh water in the bay. We use this latter measure for all quantitative estimates of flushing times. Estimates for Narragansett Bay, RI are based on the empirically derived relationship between fresh water input to the bay and flushing time²⁷. River flow data were obtained from the Rhode Island Department of Water Resources.

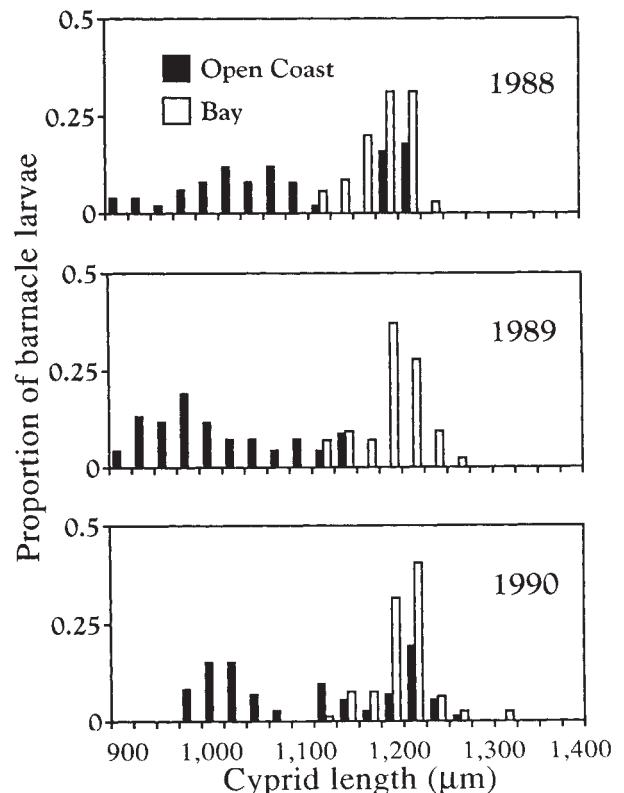


FIG. 2 Size frequency distributions for barnacle larvae collected at bay and coastal sites. Lengths were measured with an ocular micrometer on a dissecting microscope. Larvae were collected in passive tube collectors²⁴. Sample sizes ranged from 162 to 200 larvae per site. Error bars are 95% confidence limits. Note that there is little overlap in the size distributions of coastal and bay larvae in 1989, a year with a long flushing time (36 days). In both 1988 and 1990, there is a bimodal distribution of larval sizes with substantial numbers of 'bay'-size larvae at coastal sites. Each of these two years had short bay flushing times (17 and 16 days, respectively).

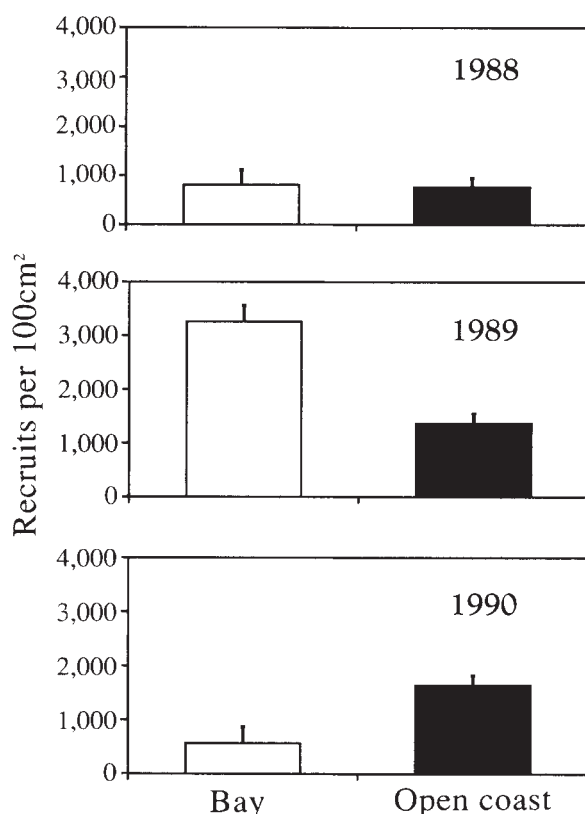


FIG. 3 Settlement rates of *Semibalanus balanoides* at bay and coastal sites. Settlement was monitored in 100-cm² quadrats ($n=16$ per site per year) at the end of the settlement season (generally mid March). Error bars are 95% confidence limits.

in exposed coastal habitats^{14,22,23}; (2) it has a short reproductive season and a single larval release, which simplifies the problems of measuring transport patterns; and (3) its larvae are accurately sampled by a new collector that integrates abundances over the entire settlement season, which makes it feasible to sample larvae at many widely separated sites simultaneously^{24,25}.

To examine the impact of larval retention on temporal variation in recruitment, we compiled a 9-year record (1982–1990) of settlement variation in Narragansett Bay, Rhode Island. The average flushing time of this bay varies substantially in response to variation in riverine input^{26,27}. There was no correlation between the interannual variation in settlement and stocks of reproductive adults. The correspondence between settlement rates and flushing times, however, is striking (Fig. 1, $R^2=0.77$, $P<0.001$). Most of the variation in nearly a decade of settlement is explained by a simple measure of the likelihood of export from the bay.

As with any conclusions based on correlation, this strong connection between settlement variation and residence times could be caused by other covarying factors. For example, here the pattern could reflect higher larval mortality in years with larger riverine inputs (and lower salinities) rather than greater losses by transport. Separating these two potential causes would typically be difficult because both larval mortality and transport would leave the same signature, declining larval densities in the bay. The solution lies in finding a marker that identifies larvae that are flushed from the bay.

Fortuitously, *S. balanoides* provides such a marker of origin, larval size. Larvae that develop within Narragansett Bay are substantially larger than larvae that develop over the continental shelf. For 3 years, we monitored larval abundance and size distributions in the water column. In a year with a long flushing time, such as 1989, there is little overlap in the size distribution of larvae collected in the bay compared with those from the open coast (Fig. 2). In years with short flushing times (1988 and 1990), there is a bimodal size distribution of larvae outside the bay. The large number of 'bay'-size larvae at coastal sites strongly supports the conclusion that the correlation of Fig. 1 is a consequence of losses due to transport. Genetic comparison of the coastal and bay larval pools further substantiate the causal role of larval transport. Transplant experiments suggest there is substantial genetic divergence of bay and coastal larval pools in retentive years, yet no divergence in years with greater transport²⁸.

Further support for the importance of retention comes from comparisons of open coast and embayed sites (Fig. 3). With long flushing times (1989), recruitment of barnacles within Narragansett Bay is significantly higher than recruitment to nearby open coast sites. In contrast, with shorter flushing times (1988 and 1990) recruitment in the bay is reduced to levels as low as or lower than those at the open coast.

The ability of retention alone to account for order of magnitude variation in settlement rates of this species poses the question of generality. Is the failure to reach favourable habitat a common source of fluctuations for other species or for other locations where patterns of local retention vary? The dynamics of herring in the North Sea¹³ as well as hake⁹ and barnacles¹² in the North Pacific all correlate with changes in transport and suggest that variable dispersal may indeed be a widespread source of recruitment variation. Beyond the obvious commercial benefits of knowing the causes of recruitment variation, the finding of a dominant role for variable dispersal would have important implications for the population biology of marine species. Variable transport, unlike variable reproductive output and variable mortality, implies that fluctuations in recruitment will covary with fluctuations in the exchange of individuals among sites. As a result, unravelling the causes of recruitment variation may be crucial to understanding the patterns of gene flow, genetic divergence and evolutionary dynamics of marine species. □

Received 17 July; accepted 19 October 1992.

- Thorson, G. *Biol. Rev.* **25**, 1–45 (1950).
- Caffey, H. M. *Ecol. Monogr.* **55**, 313–332 (1985).
- Raimondi, P. T. *Ecol. Monogr.* **60**, 283–309 (1990).
- Gaines, S. D. & Roughgarden, J. *Proc. natn. Acad. Sci. U.S.A.* **82**, 3707–3711 (1985).
- Hughes, T. P. *Ecology* **71**, 12–20 (1990).
- Sutherland, J. P. *Ecology* **71**, 955–972 (1990).
- Hjort, J. *Rapp. P.-v. Réun. Cons. perm. int. Explor. Mer* **20**, 1–228 (1914).
- Sissenwine, M. P. in *Exploitation of Marine Communities* (ed. May, R.M.) 59–94 (Springer, Berlin, 1984).
- Bailey, K. M. *Mar. Ecol. Prog. Ser.* **6**, 1–9 (1981).
- Parrish, R. H., Nelson, C. S. & Bakun, A. *Biol. Oceanogr.* **1**, 175–203 (1981).
- Kendall, M. A., Bowman, R. S., Williamson, P. & Lewis, J. R. *J. mar. biol. Ass. UK* **65**, 1009–1030 (1985).
- Roughgarden, J., Gaines, S. D. & Possingham, H. *Science* **241**, 1460–1466 (1988).
- Bartsch, J. et al. *Nature* **340**, 632–636 (1989).
- Hawkins, S. J. & Hartnoll, R. G. *J. exp. mar. Biol. Ecol.* **62**, 271–283 (1982).
- Kendall, M. A., Bauman, R. S., Williamson, P. & Lewis, J. R. *Neth. J. Sea Res.* **16**, 119–126 (1982).

- Zeldis, J. R. & Jillett, J. B. *J. Plankton Res.* **4**, 839–857 (1982).
- Shanks, A. L. *Biol. Bull.* **170**, 429–440 (1986).
- Geyer, W. R. & Signell, R. P. *Estuaries* **15**, 97–108 (1992).
- Ketchum, B. H. *Ecology* **35**, 191–200 (1954).
- Sulkin, S. D. & VanHeukelen, W. V. in *Estuarine Comparisons* (ed. Kennedy, V. S.) 459–475 (Academic, New York, 1982).
- Morgan, S. G. *Ecology* **71**, 1639–1652 (1990).
- Menge, B. A. *Ecol. Monogr.* **46**, 355–393 (1976).
- Southward, A. J. *J. mar. biol. Ass. UK* **71**, 495–513 (1991).
- Yund, P. O., Gaines, S. D. & Bertness, M. D. *Limnol. Oceanogr.* **36**, 1167–1177 (1991).
- Bertness, M. D., Gaines, S. D., Stephens, E. G. & Yund, P. O. *J. exp. mar. Biol. Ecol.* **156**, 199–215 (1992).
- Bolin, B. & Rodhe, H. *Tellus* **25**, 58–62 (1973).
- Pilson, M. E. Q. *Estuaries* **8**, 2–14 (1985).
- Bertness, M. D. & Gaines, S. D. *Evolution* (in the press).

ACKNOWLEDGEMENTS. We thank D. Bermudez, E. Sanford, E. Stephens and P. Yund for field assistance and S. Hawkins for comments on the manuscript. This work was supported by the National Science Foundation and a Brown University Research Support Grant to S.D.G.

Communal nesting patterns in mice implicate MHC genes in kin recognition

C. Jo Manning*†, Edward K. Wakeland* & Wayne K. Potts*

* Center for Mammalian Genetics and Department of Pathology and Laboratory Medicine, University of Florida, Gainesville, Florida 32610, USA

† Department of Psychology, University of Washington, Seattle, Washington 98195, USA

HOUSE mice (*Mus musculus domesticus*) form communal nests and appear to nurse each other's pups indiscriminately. Communal nesting probably functions to reduce infanticide¹, but it also makes females vulnerable to exploitation if nursing partners fail to provide their fair share of care. Kinship theory predicts that females will preferentially form communal nests with relatives to minimize exploitation and further increase inclusive fitness²⁻⁴. Here we provide evidence from seminatural populations that females prefer communal nesting partners that share allelic forms of major histocompatibility complex genes. Such behaviour would lead to the selection of close relatives as communal nesting partners⁵⁻⁷. Although criteria for the demonstration of kin recognition are currently embroiled in controversy^{8,9}, this is the first vertebrate study to meet Grafen's restrictive requirements^{8,10}: discrimination is based on genetic similarity at highly polymorphic loci, incidental correlations due to relatedness are experimentally controlled, and strong reasons exist for expecting the assayed behaviour to be kin-selected.

Major histocompatibility complex (MHC) gene products are cell-surface molecules that function during immune recognition¹¹, but they also fulfill the two major requirements for genetically based kin recognition^{6,7,12-17}. These polymorphic loci strongly influence individual odour profiles¹⁸, providing a mechanism for phenotypic discrimination, and their unprecedented genetic diversity (often over 100 alleles per locus) allows useful resolution of genetic relationships⁵⁻⁷. Consistent with this MHC-based kin recognition hypothesis are observations of MHC-disassortative mating preferences in house mice¹⁹⁻²¹, which may function to avoid inbreeding^{13,15-17,21}.

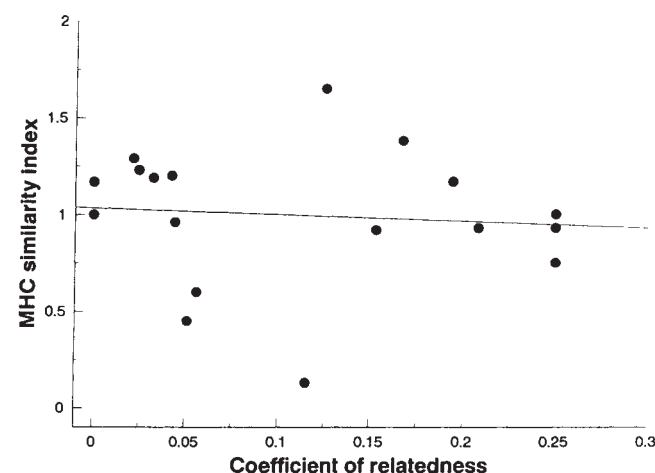


FIG. 1 Correlation between MHC similarity and relatedness. Observations of nests containing non-sibs from populations E and G-J show no correlation ($r^2 = -0.008$) between relatedness (r between communally nesting female and the mean of all nests available) and the MHC similarity index of nests available. Coefficients of relatedness (r) were calculated for all potential communal nesters by path analysis, using colony records.

To test the hypothesis that MHC plays a role in the choice of communal nesting partners, we studied six seminatural populations of *Mus*. All mice were half-wild derived but carried combinations of four MHC haplotypes derived from inbred strains (b from C57BL/6, d from BALB/c, k from B10.BR and q from DBA/1) (see ref. 21 for breeding design). In populations C and G-J (Table 1), founders had been reared to 90 days of age in cages with their same sex sibs. Founding populations had one to three sister groups of two to five individuals. Except for these cagemate sister groups, all founders were strangers. Population E consisted of offspring born in the enclosure and left to found a new population after removal of their parents and younger sibs. Table 1 provides the numbers of single and communal nests, founders and pups born for each population.

We identified all communal nesting options for each mother on the day she gave birth and calculated the MHC similarity of the mother to each of the dams in each prospective nest (see legend to Table 2). MHC similarity scores for nests chosen and nests available were compared to test the null hypothesis that nesting preferences were random with respect to MHC.

Because MHC-based kin recognition by phenotype matching could operate in several ways, eight models were created to test the range of possibilities (Table 2). These models include variations on the matching referent or template (self, nestmates or parents), whether matching is to genotype or haplotype, and whether mean or maximum similarity is used when nests consist of more than two females (see Table 2 legend). An additional variable had to be considered because sisters reared as cagemates preferentially nested together. We controlled for this variable in two ways. First, a separate analysis was done on the subset (50%) of communal nests not involving cagemates. Second, in an analysis using the entire data set, we weighted the null model (Table 2 legend) to control for the preference for cagemate sisters. This weighted data set still contained MHC similarity information when more than one nest option contained a cagemate sister or when cagemate sisters were not chosen. Table 2 shows that MHC similarity of the chosen nest was significantly greater ($P < 0.05$) than would have been achieved by random chance for all eight models when the entire data set was analysed ($n = 50$). When communal nests not involving cagemates were analysed separately ($n = 25$), six of eight models showed significant differences ($P < 0.05$) and the other two showed similar trends.

These results are consistent with the hypothesis that MHC similarity is involved in the choice of communal nesting partners. However, partners may be chosen on the basis of loci unlinked but correlated with MHC loci. Such correlations will normally exist among relatives so that preferences based on discrimination of other genes or on behavioural similarities among kin⁹ could appear to be MHC-mediated. But owing to our breeding design, correlations between relatedness and MHC similarity were eliminated for non-sibs in all populations except population C (see ref. 21). Figure 1 confirms that the correlation between MHC and relatedness was abolished ($r^2 = -0.008$). In the subset of nests involving no correlation between MHC and relatedness, there is still a significant trend for MHC similarity among nesting partners (Table 2, first model; $n = 18$). These findings are inconsistent with the hypothesis that preferences based on genes unlinked to, but correlated with, MHC could explain the high MHC similarity among nesting partners. Furthermore, this data subset reveals that the coefficient of relatedness (r) for nests chosen (mean $r = 0.099$) and nests available (mean $r = 0.110$) do not differ ($n = 18$, $P = 0.46$), demonstrating that other (non-MHC) indicators of relatedness are not being used to choose communal nesting partners in these populations.

These data suggest that female house mice use an MHC-based genetic matching system in choosing communal nesting partners. In our view, kin recognition is the most likely interpretation of these findings because of expectations that kin will be preferred as nesting partners^{3,4} and because this is the only hypothesis

TABLE 1 A comparison of communal and single-mother nests in the six study populations

Population	Dates	Founders		Single-mother nests		Communal nests		
		♂	♀	Number of nests	Pups born	Number of nests	Number of litters	Pups born
C	4/8-6/12/88	8	16	17	103	4	11	66
E	23/9-6/12/88	12	19	7	44	6	20	118
G	5/5-27/7/89	8	16	7	37	5	15	91
H	24/4-7/7/89	8	16	8	40	4	8	47
I	31/7-9/11/89	8	16	5	35	3	13	70
J	31/7-9/11/89	8	16	14	87	2	3	10
Totals		52	99	58	346	24	70	402

Study populations: Our 9.8 × 4.9 m enclosures were of sufficient size and complexity to allow the formation of several male territories and social units. Unique ear punch combinations allowed identification of all individuals through binoculars. Daytime nest checks and 1–2h of behavioural observations at dusk were made 5–7 times a week. **Communal nests:** Most communal nests were formed by a mother giving birth in an already existing nest but some were combined after separate births. We quantified properties of these nests, such as the number of litters or dams already in a nest and the age of the pups. A comparison of the means for all nests available to the nest that was chosen revealed no difference for number of litters (1.26 versus 1.25 litters, $P > 0.90$) or nursing dams per nest (1.42 versus 1.48 dams, $P > 0.25$), but the age of pups in nests chosen was significantly less than the mean age of pups in all nests available (5.5 versus 6.2 days, $P = 0.009$). Although we observed 70 females who communally nested, only 50 of these females had more than one communal nest option (mean = 3.6, range 2–6). Although similar numbers of pups were born in communal and single-mother nests, these two nest types were not equally preferred. Most single-mother nests occurred when communal nesting options were not available. Ninety per cent of mothers ultimately nested communally.

TABLE 2 A comparison of MHC similarity between nests chosen and nests available

Model	N	MHC similarity to nest chosen (±s.e.)	MHC similarity to mean of nests available (±s.e.)	P
Match to self				
haplotype, mean MHC similarity	50	1.33 ± 0.08	1.19 ± 0.07	0.014*
(without cagemates)	25	1.18 ± 0.10	0.95 ± 0.07	0.036*
(without population C or cagemates)	18	1.20 ± 0.11	1.00 ± 0.08	0.50*
Match to self haplotype, maximum MHC similarity	50	1.45 ± 0.08	1.33 ± 0.07	0.040*
(without cagemates)	25	1.30 ± 0.09	1.06 ± 0.08	0.022*
Match to nestmate				
haplotype, mean MHC similarity	50	1.71 ± 0.06	1.55 ± 0.06	0.002*
(without cagemates)	25	1.49 ± 0.11	1.27 ± 0.09	0.021*
Match to nestmate				
haplotype, maximum MHC similarity	50	1.77 ± 0.07	1.66 ± 0.06	0.018*
(without cagemates)	25	1.58 ± 0.10	1.35 ± 0.10	0.022*
Match to nestmate				
genotype, mean MHC similarity	50	1.49 ± 0.11	1.33 ± 0.09	0.025*
(without cagemates)	25	1.09 ± 0.19	0.93 ± 0.12	0.148
Match to parental				
haplotype, mean MHC similarity	50	1.38 ± 0.08	1.25 ± 0.06	0.022*
(without cagemates)	25	1.21 ± 0.11	1.04 ± 0.06	0.064
Match to parental				
haplotype, maximum MHC similarity	50	1.54 ± 0.07	1.41 ± 0.06	0.014*
(without cagemates)	25	1.40 ± 0.11	1.17 ± 0.09	0.027*
Match to parental				
genotype, mean MHC similarity	50	0.97 ± 0.13	0.78 ± 0.09	0.012*
(without cagemates)	25	0.78 ± 0.18	0.45 ± 0.07	0.013*

Models: Work by Yamazaki *et al.*²⁵ suggests that mice use parents as referents for mate choice. But because fathers are not generally present in the nest in nature and because communal nesting may cause confusion about maternity, the efficacy of such a system is doubtful. Accordingly, we have modelled all reasonable possibilities. Match to self assumes self is the odour referent to whom matching is done. Match to nestmates assumes all nestmates and both parents can serve as referents. Match to parents

assumes both parents are referents. The set of loci that comprise the MHC are tightly linked, allowing the haplotype to be treated as the unit of inheritance. Matching was analysed at both the genotypic and haplotypic levels. A genotypic match requires identical genotypes. Haplotype matching can be complete (identical genotypes) or partial (matching one haplotype). Mean MHC similarity score for each nest is the mean of the scores for all dams in the nest. Mean MHC similarity index score for all nests available is the average of all nest means. Maximum MHC similarity scores are derived by taking only the score of the female in the nest with the highest similarity score. For mean of nests available, maximum scores of individual nests are averaged. Mean MHC similarity scores provide a way for a female to evaluate her relatedness to the nest as a whole. Thus a nest with non-relatives in addition to relatives is proportionately devalued. Maximum scores assume that a female choosing a nest only cares about the one individual in that nest that is most MHC similar. **MHC similarity index:** MHC similarity for matching to self models is scored as 0 if all four haplotypes of two individuals are different, 1 for two heterozygotes with one haplotype in common (such as bd and bk), 1.5 for a heterozygote with one haplotype matching both haplotypes of a homozygote (such as bd and bb), and 2 for either identical homozygotes (bb and bb) or heterozygotes (bd and bd). Identical homozygotes or heterozygotes (score of 2) would also be scored as a genotype match. When the matching referent is nestmates or parents there are several complicated ways a partial match can occur. Accordingly, the MHC similarity index is simplified as 0 = no match, 1 = partial match, 2 = complete or genotype match. **Nesting options:** A nest was considered a communal nesting option if it contained pups no more than 17 days of age or if the dam(s) had lost a litter within three days. These pupless dams always demonstrated normal communal nesting behaviour toward the new litter.

Weighting of the models: The null model assumes that females will choose randomly with respect to MHC. But because females chose a nest containing a former cagemate 92% of the time one was available, nest choice was not random with respect to former cagemates. To control for this we conditioned the model on the probability that a female will choose to nest with a cagemate and nesting options were weighted in the following way. For the five populations containing only familiar sib cagemates or unfamiliar non-cagemates, an estimate of the likelihood that a female would choose a cagemate nest whenever both a cagemate (type *i*) and a non-cagemate (type *j*) option were available was calculated as $P_i = X_i / N_{i+j}$ where X_i is the number of times a cagemate nest (type *i*) was chosen and N_{i+j} is the number of times both cagemate nest and non-cagemate nests were available. Mean MHC similarity of all nests available to each dam on the day she gave birth was computed as $M_d = \sum_{aj} P_{aj} M_{dj}$ where M_{dj} denotes MHC similarity between dam *d* and nest options *j*. Although the weighting process eliminated much of the power of the data involving cagemates, additional information is gained because females tended to choose a cagemate sister with a higher MHC similarity index when more than one such nest was available. Also, on the rare occasions when females did not choose a nest containing a cagemate, they tended to choose nests with high MHC similarity indices. **Statistic:** A single sample *t* test was used to test the null hypothesis that the difference between MHC similarity to nest chosen and the mean MHC similarity of nests available was zero.

* $P < 0.05$.

predicting both MHC-assortative preferences during cooperative behaviour and MHC-disassortative preferences during mating^{19–21}. (A putative kin recognition system in the tunicate *Botryllus schlosseri* shows remarkable similarities, where a highly polymorphic gene influences preferences during both mating and cooperative behaviour^{14,22,23}.) But these findings are also consistent with the hypothesis that females with MHC-familiar odours are preferred because long-term associations between 'group members' are favoured^{9,24}. If so, the observed preferences for unfamiliar but MHC-similar partners would represent a case of mistaken 'group' identity. In either case, to our knowledge this is the first demonstration in a vertebrate of specific genes used to target individuals for differential treatment during cooperative behaviour. □

Received 11 May; accepted 1 October 1992.

1. Manning, C. J., Dewsbury, D. A., Wakeland, E. K. & Potts, W. K. *Anim. Behav.* (in the press).
2. Hamilton, W. D. *J. theor. Biol.* **7**, 1–52 (1964).
3. Wilkinson, G. S. & Baker, A. E. M. *Ethology* **77**, 103–114 (1988).
4. Packer, C., Lewis, S. & Pusey, A. *Anim. Behav.* **43**, 265–281 (1992).
5. Getz, W. M. *J. theor. Biol.* **92**, 209–226 (1981).
6. Beecher, M. D. *Am. Zool.* **22**, 477–490 (1982).
7. Lacy, R. C. & Sherman, P. W. *Am. Nat.* **121**, 489–512 (1983).

8. Grafen, A. *Anim. Behav.* **39**, 42–54 (1990).
9. Barnard, C. J., Hurst, J. L. & Aldous, P. *Biol. Rev.* **66**, 377–430 (1991).
10. Barnard, C. *Trends Ecol. Evol.* **310**, 312 (1991).
11. Klein, J. *Natural History of the Major Histocompatibility Complex* (Wiley, New York, 1986).
12. Thomas, L. in *Fourth International Convocation of Immunology* (eds Neter, E. & Milgrom, F.) 2–11 (Karger, Basel, 1975).
13. Brown, J. L. in *Ethical Questions in Brain and Behavior* (ed. Pfaff, D. W.) 111–124 (Springer, New York, 1983).
14. Grosberg, R. K. & Quinn, J. F. *Nature* **322**, 456–459 (1986).
15. Partridge, L. *Philos. Trans. R. Soc. B* **319**, 525–539 (1988).
16. Uyenoyama, M. K. in *The Evolution of Sex* (eds Michod, R. E. & Levin, B. R.) 212–232 (Sinauer, Sunderland, MA, 1988).
17. Potts, W. K. & Wakeland, E. K. *Trends Ecol. Evol.* **5**, 181–187 (1990).
18. Beauchamp, G. K., Yamazaki, K., Duncan, H., Bard, J. & Boyse, E. A. in *Chemical Signals in Vertebrates V* (eds MacDonald, D. W., Muller-Schwarze, D. & Natynczuk, S. E.) 244–254 (Oxford Science Publications, Oxford, 1990).
19. Beauchamp, G. K., Yamazaki, K., Bard, J. & Boyse, E. A. *Behav. Genet.* **18**, 537–547 (1988).
20. Egid, K. & Brown, J. L. *Anim. Behav.* **38**, 548–549 (1989).
21. Potts, W. K., Manning, C. J. & Wakeland, E. K. *Nature* **352**, 619–621 (1991).
22. Oka, H. in *Profiles of Japanese Science and Scientists 1970* (ed. Yukawa, H.) 196–206 (Kodansha, Tokyo, 1970).
23. Scofield, V. L., Schlumberger, J. M., West, L. A. & Weissman, I. L. *Nature* **295**, 499–502 (1982).
24. Hurst, J. L. & Barnard, C. J. *Behav. Ecol.* **3**, 196–206 (1992).
25. Yamazaki, K. *et al. Science* **240**, 1331–1332 (1988).

ACKNOWLEDGEMENTS. We thank J. Longmate for advice on data analysis and statistics, D. Barash, C. Barnard, M. Beecher, A. Blaustein, J. Brown, D. Dewsbury, A. Grafen and R. Grosberg for advice on the manuscript, undergraduates and technicians who have worked on this project and D. Dewsbury for hosting C.J.M. at the University of Florida. This work was supported by a grant from the NIH and by an NSF graduate fellowship to C.J.M.

The perception of heading during eye movements

Constance S. Royden*, Martin S. Banks†
& James A. Crowell

School of Optometry and Department of Psychology,
University of California at Berkeley, Berkeley, California 94720, USA

* Present address: MIT Media Laboratory, Massachusetts Institute of
Technology, Cambridge, Massachusetts 02139, USA

† To whom correspondence should be addressed

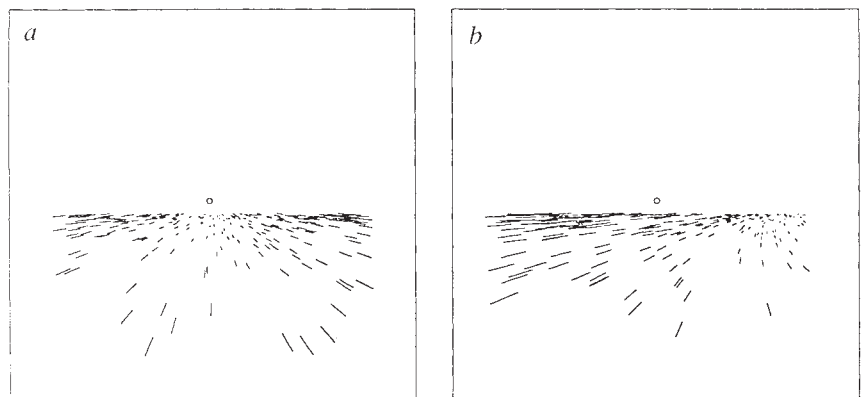
WHEN a person walks through a rigid environment while holding eyes and head fixed, the pattern of retinal motion flows radially away from a point, the focus of expansion (Fig. 1a)^{1,2}. Under such conditions of translation, heading corresponds to the focus of expansion and people identify it readily³. But when making an eye/head movement to track an object off to the side, retinal motion is no longer radial (Fig. 1b)⁴. Heading perception in such situations has been modelled in two ways. Extra-retinal models monitor the velocity of rotational movements through proprioceptive or efference information from the extraocular and neck muscles and use that information to discount rotation effects⁵. Retinal-image models determine (and eliminate) rotational components from the retinal image alone^{6–12}. These models have been tested^{13,14} by measuring heading perception under two conditions. First, obser-

vers judged heading while tracking a point on a simulated ground plane. Second, they fixated a stationary point and the flow field simulated the effects of a tracking eye movement. Extra-retinal models⁵ predict poorer performance in the simulated condition because the eyes do not move. Retinal-image models^{6–12} predict no difference in performance because the two conditions produce identical patterns of retinal motion. Warren and Hannon^{13,14} observed similar performance and concluded that people do not require extra-retinal information to judge heading with eye/head movements present, but they used extremely slow tracking eye movements of 0.2–1.2 deg s⁻¹; a moving observer frequently tracks objects at much higher rates (L. Stark, personal communication). Here we examine heading judgements at higher, more typical eye movement velocities and find that people require extra-retinal information about eye position¹⁵ to perceive heading accurately under many viewing conditions.

Experiment 1 reproduced the conditions of Warren and Hannon¹³ except we used constant, faster rotation rates (actual or simulated) from 0 to 5 deg s⁻¹ and a vertical axis of rotation. In addition, the fixation point was positioned slightly above the horizon and moved independently of the ground plane. At the end of a stimulus presentation, seven vertical lines appeared and the observers indicated the one that corresponded most closely to the perceived heading. As in the experiments of Warren and Hannon^{13,14}, real and simulated eye movement conditions produced identical patterns of retinal image motion, so retinal-image models predict similar performance in the two conditions.

Figure 2 shows that both observers responded very differently

FIG. 1 Optical flow fields for an observer moving across a ground plane. *a*, Flow field for a translational movement; the observer has moved forward while holding the eye and head fixed. The circle marks the focus of expansion, which corresponds to the observer's direction of motion. *b*, Flow field for translation plus rotation; the observer has again moved forward while tracking an object moving from left to right. As before, the circle marks the observer's heading. In the real eye movement condition of experiment 1, the flow field on the display screen resembled the one in *a* and the flow field on the retina resembled the one in *b*; in the simulated eye movement condition, the flow fields on the display screen and retina resembled the one in *b*.



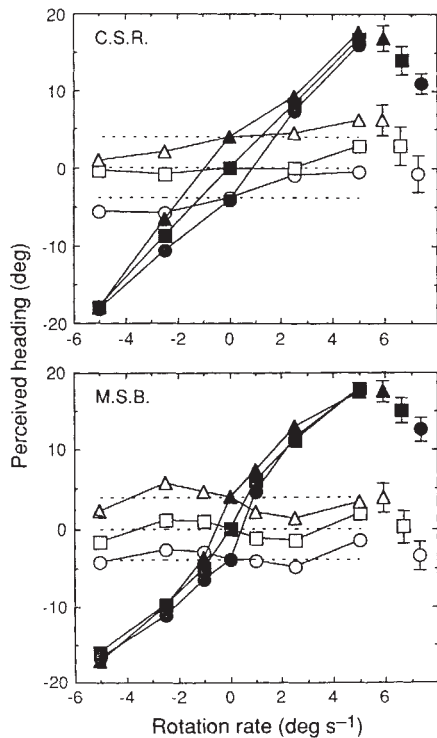


FIG. 2 Results of experiment 1: heading judgements for motion across a ground plane with real or simulated eye movements. Dot motions corresponded to the optical flow produced when an observer walks across a flat surface at a speed of 190 cm s^{-1} . Simulated eye height was 160 cm. The ground plane was truncated at a distance of 3,730 cm. Dots were distributed randomly on the plane with an average density of 0.6 dots m^{-2} ; roughly 220 dots were visible at the beginning of a trial. The stimulus subtended $30 \times 30 \text{ deg}$ at the 30-cm viewing distance. Viewing was monocular. The fixation point was 2.5 deg above the truncated horizon. In the real eye movement condition, observers C.S.R. and M.S.B. tracked a point moving horizontally at 0, ± 2.5 or $\pm 5 \text{ deg s}^{-1}$. In the simulated eye movement condition, observers fixated a stationary point and the flow field simulated the effects of horizontal eye movements at 0, ± 2.5 and $\pm 5 \text{ deg s}^{-1}$. In addition, observer M.S.B. was tested for rotation rates of $\pm 1 \text{ deg s}^{-1}$ in both conditions. At the beginning of the motion sequence, the fixation point (which actually began moving 200 ms earlier) was always at the center of the display, so at that instant the heading was always towards or 4 deg to the left or right of fixation. These eccentricities were identical across the real and simulated conditions. At the end of each, 1,250-ms trial, observers indicated which of 7 target lines, equally spaced 4 deg apart, was closest to the perceived heading; the target lines were 2 deg apart when the rotation rate was $\pm 1 \text{ deg s}^{-1}$. Each condition was presented 20 times. The two panels show the results separately for two observers; others yielded similar data. Open and filled symbols represent responses in the real and simulated eye movement conditions, respectively. Circles, squares and triangles represent the responses for headings of -4 , 0 and 4 deg . Error bars on the right show twice the average standard deviation for each heading.

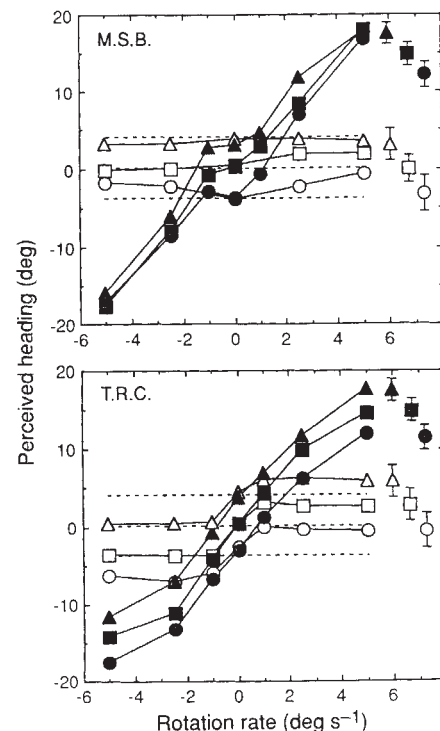
in the two conditions: they judged heading accurately in the real eye movement condition and very inaccurately in the simulated condition. When they moved their eyes, the average errors were 1.5 and 1.9 deg for rotation rates of 2.5 and 5 deg s^{-1} , respectively; when they did not, the errors increased to 9.8 and 17.3 deg. The results suggest that humans require proprioceptive or efferent information from the extra-ocular muscles to judge heading accurately in the presence of rotations. Most retinal-image models^{6-8,11} fail to predict the large inaccuracies in heading judgements seen in the simulated condition.

The retinal-image models require depth variation in the scene in order to separate translational and rotational flow⁶⁻¹². Con-

sequently, they cannot find a unique solution for translation relative to a frontoparallel plane. To see if human observers have similar difficulties, we repeated experiment 1 using a frontoparallel plane instead of a ground plane. In experiment 2, both observers again made accurate heading judgements in the real eye movement condition (average errors were 1.1 and 1.5 deg at 2.5 and 5 deg s^{-1} , respectively) and inaccurate ones in the simulated condition (average errors were 9.1 and 16.8 deg). Thus, when extra-retinal information about eye position is available, observers make accurate heading judgements in a situation in which retinal-image models⁶⁻¹² do not.

Because our results conflict with those of Warren and

FIG. 3 Results of experiment 3: heading judgements for observer motion (observers M.S.B. and T.R.C.) through a rigid, 3-dimensional cloud of dots with real or simulated eye movements. The motion of the dots in the display simulated translation at 50 cm s^{-1} through a cloud with dots at distances of 0–3,730 cm. Roughly 615 dots were visible at the beginning of the trial. The fixation point was a member of the rigid cloud and was positioned 5 deg to the left or right of the heading at the beginning of the motion sequence. Placing it at distances of 107, 310, 162 and 112 cm produced average rotation rates 0, 1, 2.5 and 5 deg s^{-1} , respectively. All other display and procedural parameters are identical to those in experiments 1 and 2. The open and filled symbols represent heading judgements in the real and simulated eye movement conditions, respectively. Circles, squares and triangles represent judgements for headings of -4 , 0 and 4 deg . The top, middle and bottom horizontal dotted lines show the true headings of -4 , 0 and 4 deg , respectively.



Hannon^{13,14}, we performed an exact replication of their experiment. As they reported, observers made reasonably accurate judgements independently of whether actual or simulated eye movements generated the rotational flow. Thus, with very slow rotations and/or with the tracked object fixed to the surface, observers can judge heading accurately with simulated eye movements.

Experiment 3 tested which of these two differences in the experiments was critical. The displays simulated translation through a rigid, three-dimensional cloud of dots. The fixation point was a member of the cloud, but we varied rotation rate by placing it at different simulated distances. Figure 3 shows that heading judgements were accurate in the real eye movement condition and poor in the simulated condition. Thus, rotation rate rather than the attachment of the fixation point to the simulated three-dimensional object appears to be the critical variable. Interestingly, judgements in the simulated condition were fairly accurate at 1 deg s^{-1} so observers can judge heading well from retinal image information alone when the simulated rotation is slow^{13,14}.

Although these experiments were done in a dark room, the edge of the display screen was just visible. Perhaps observers did not interpret the rotational flow in the simulated condition as a consequence of eye movements because the screen edge

did not move relative to the fixation point. We repeated experiments 1 and 3 under conditions in which only the dots were visible and got the same results. Therefore this possible artefact cannot explain the erroneous judgements in the simulated eye movement condition.

Observers badly misperceived their heading in the simulated eye movement conditions. When presented the frontoparallel plane of experiment 2, observers thought they were heading towards the position of zero flow in the display. With the ground plane of experiment 1, they thought they were moving along a curvilinear path in the direction of the simulated eye movement. In neither case did they perceive the specified linear motion plus horizontal eye movement. The erroneous percept in the simulated condition of experiment 1 reflects the fact that the pattern of retinal motion closely resembles the pattern created by motion on a curvilinear path¹⁶. Presumably, such misperceptions do not occur in the real world with observer-initiated locomotion because extra-retinal information helps distinguish linear motion plus eye/head rotation from curvilinear motion. We have shown that humans require extra-retinal information about eye position to perceive heading accurately in the presence of rotation rates greater than $1 \text{ deg}^{-1} \text{ s}$. These rates occur when an observer fixates a nearby object that is not along his or her heading, or fixates a moving object. □

Received 17 August; accepted 13 October 1992.

1. Gibson, J. J. *The Perception of the Visual World* (Houghton Mifflin, Boston, 1950).
2. Gibson, J. J. *The Senses Considered as Perceptual Systems* (Houghton Mifflin, Boston, 1966).
3. Warren, W. H., Morris, M. W. & Kalish, M. J. *exp. Psych.* **14**, 646-660 (1988).
4. Regan, D. & Beverley, K. I. *Science* **215**, 194-196 (1982).
5. von Holst, E. *Anim. Behav.* **2**, 89-94 (1954).
6. Longuet-Higgins, H. C. & Prazdny, K. *Proc. R. Soc. Lond. B* **208**, 385-397 (1980).
7. Rieger, J. H. & Lawton, D. T. *J. opt. Soc. Am. A* **2**, 354-360 (1985).
8. Heeger, D. J. & Jepson, A. D. *int. J. comp. Vis.* (in the press).
9. Bruss, A. R. & Horn, B. K. P. *Comp. Vis. Graph Im. Proc.* **21**, 3-20 (1983).
10. Waxman, A. M. & Ullman, S. *Int. J. robot. Res.* **4**, 72-94 (1985).
11. Droulez, J. & Cornilleau-Peres, V. *Biol. Cybern.* **62**, 211-224 (1990).

12. Longuet-Higgins, H. C. *Nature* **293**, 133-135 (1981).
13. Warren, W. H. & Hannon, D. J. *Nature* **336**, 162-163 (1988).
14. Warren, W. H. & Hannon, D. J. *J. opt. Soc. Am. A* **6**, 160-169 (1990).
15. Matin, L. in *Symp. Study of Motion Perception* (eds Wagenaar, W. A., Wertheim, A. H. & Leibowitz, H. W.) (Plenum, New York, 1981).
16. Warren, W. H., Blacknell, A. W., Kurtz, K. J., Hatsopoulos, N. G. & Kalish, M. L. *Biol. Cybern.* **65**, 311-320 (1991).

ACKNOWLEDGEMENTS. This research was supported by an NIH research grant and a NASA Ames Research Center grant (M.S.B.). C.S.R. is a Burroughs Wellcome Fund Fellow of the Life Sciences Research Foundation. The authors acknowledge the helpful comments of L. K. Cormack, S. B. Stevenson and W. H. Warren.

Three-dimensional illusory contours and surfaces

George J. Carman* & Leslie Welch†

* The Salk Institute VCL, PO Box 85800, San Diego, California 92186, USA
† University of California, San Diego, California 92093, USA

UNDER general viewing conditions, objects are often partially camouflaged, obscured or occluded, thereby limiting information about their three-dimensional position, orientation and shape to incomplete and variable image cues. When presented with such partial cues, observers report perceiving 'illusory' contours and surfaces (forms) in regions having no physical image contrast¹⁻⁷. Here we report that three-dimensional illusory forms share three fundamental properties with 'real' forms: (1) the same forms are perceived using either stereo or motion parallax cues (cue invariance^{8,9}); (2) they retain their shape over changes in position and orientation relative to an observer (view stability¹⁰); and (3) they can take the shape of general contours and surfaces in three dimensions¹¹ (morphic generality). We hypothesize that illusory contours and surfaces are manifestations of a previously unnoticed visual process which constructs a representation of three-dimensional position, orientation and shape of objects from available image cues.

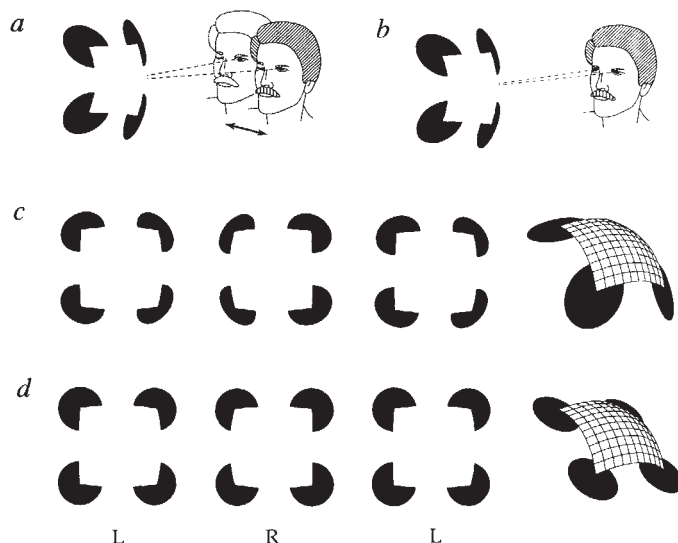
We stimulated natural viewing of a set of inducing discs which were partially occluded by an 'invisible' three-dimensional surface having no contrast with the surround^{4,12} (Fig. 1). Relative depth information could be presented using either or both motion and stereo cues: motion parallax was presented by simulating motion of the observer relative to the inducing discs

(Fig. 1a), whereas stereo parallax was presented by stimulating binocular viewing of the inducing discs (Fig. 1b). When these cues specified no relative depth, only coplanar illusory forms were induced. But when either cue presented a depth gradient or discontinuity, three-dimensional illusory contours and surfaces could be induced^{4,12}. Observers consistently reported that these illusory contours and surfaces had the same position, orientation and shape regardless of whether they were induced by motion, or stereo, or both cues together (see Fig. 2 legend). Furthermore, the same illusory forms could be induced by discs having various underlying three-dimensional shapes (Fig. 1c, d). Although earlier results indicated that motion cues¹³⁻¹⁵ or stereo cues^{12,16,17} were each capable of inducing illusory contours and figures, ours are the first findings demonstrating that the same three-dimensional illusory forms can be induced by different depth cues. Such cue invariance indicates that illusory forms can participate in the perception of three-dimensional objects despite changes in available cues occurring under general viewing conditions.

We were also able to simulate changes in the viewing of illusory contours and surfaces by presenting discs which were partially occluded by an 'invisible' surface whose position and orientation with respect to the observer we could manipulate (Fig. 2). Earlier studies had established that illusory contours could take a variety of orientations in the plane⁵ and in depth¹⁶, and that illusory surfaces could be induced at different depths^{12,16,17,31}. We found that we were able to induce the same three-dimensional illusory contours and surfaces as if they were being viewed from a variety of different perspectives, resulting in substantial changes in the three-dimensional position and orientation of the illusory forms. Although the size of the illusory form was scaled by changes in depth (Fig. 2a, b), and the aspect

FIG. 1 Cue invariance of three-dimensional illusory contours and surfaces. *a, b*, Schematic of simulated natural viewing of four partially occluded discs presenting *a*, monocular motion parallax cues, or *b*, binocular stereo parallax cues to relative depth. Either or both cues could induce illusory contours and surfaces having the same three-dimensional position, orientation and shape. Examples of such three-dimensional illusory forms are demonstrated using stereo cues in *c, d*. Each row presents stereo pairs containing four discs which induce three-dimensional illusory contours and surfaces when fused (first and second column for uncrossed viewing; second and third column for crossed viewing; L and R designate left and right eye views). Each row also presents a perspective view depicting the percept obtained upon stereo fusion (fourth column: inducing discs, solid; illusory surface, grid). The same illusory contours and surfaces could be induced by discs located on surfaces having either the same or different underlying shape. *c*, Inducing discs located on the same surface provide intrinsic depth cues which can be interpreted in terms of classical stereo or motion parallax mechanisms^{12,18}. *d*, Inducing discs located on a different surface provide extrinsic depth cues which require a non-classical mechanism attributing parallax presented by the occluded discs to a separate, illusory surface rather than to the discs themselves^{12,18}.

METHODS. Inducing stimuli were created on a Silicon Graphics Iris 4D20G graphics workstation, which displayed dynamic, stereo pairs of perspective images on a 1,280 by 1,024 resolution screen at 30 Hz, and were viewed with a mirror haploscope with optical path length of 57.3 cm. Motion parallax simulated 8° s^{-1} rotation of the observer relative to the inducing discs. Stereo parallax simulated differential perspective appropriate for each eye's view of the inducing discs.



presented to the observer was altered dramatically with changes in orientation (Fig. 2*b, c*), observers consistently reported that the shape of three-dimensional illusory contours and surfaces remained constant over such changes in view. Such view stability strongly suggests that illusory forms can participate in the perception of the shape of three-dimensional objects, despite changes in the viewpoint of an observer relative to an object that occur under general viewing conditions.

We were also able to induce illusory contours and surfaces having any of the four possible elementary three-dimensional shapes. All smooth, continuous surfaces can be decomposed on the basis of their local surface curvature into regions which are termed planar, parabolic, elliptic and hyperbolic¹¹. We induced illusory contours and surfaces having each of these four elementary shapes by presenting discs which were partially occluded by an 'invisible' surface whose curvature we could manipulate (Fig. 3). Observers consistently reported perceiving illusory contours and surfaces corresponding to each of these elementary shapes, thereby demonstrating the morphic generality essential for illusory forms to participate in the detection and discrimina-

tion of general three-dimensional objects. Although the occluded edges of the discs are consistent with the perceived forms^{12,16,18,31}, they do not provide sufficient information to specify the local position, orientation, and shape (curvature) of the surfaces inside their boundary contours. Thus, the perception of three-dimensional illusory surfaces must be a consequence of some process of interpolation, completion or inference which is capable of constructing a representation of smooth three-dimensional surfaces from such partial cues.

We have demonstrated that the perception of three-dimensional illusory contours and surfaces is characterized by cue invariance, view stability and morphic generality, as is the perception of real forms. The conjunction of all three properties strongly supports our hypothesis that perception of both real and illusory forms results from a visual process which constructs a representation of three-dimensional geometry from various partial cues^{4,9,19-21} (Fig. 4). We term this process morphopsis (form-vision) in order to distinguish it from earlier visual processes, such as kineopsis and stereopsis, to which surface interpolation has been ascribed²¹⁻²³. Our finding that this process

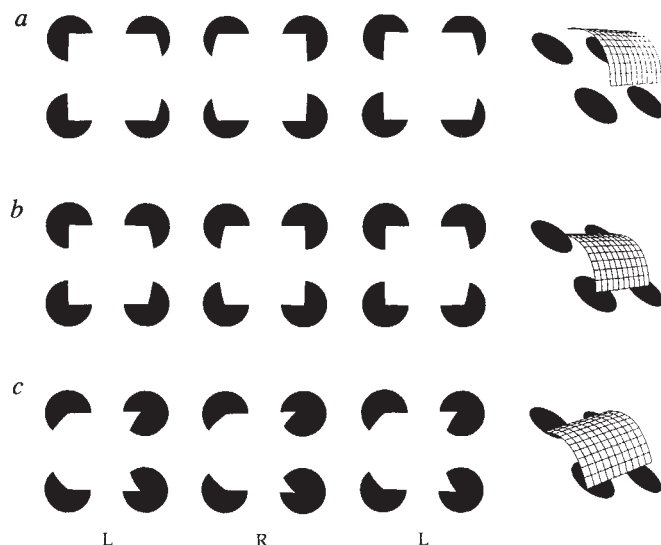


FIG. 2 View stability over changes in position and orientation of three-dimensional illusory contours and surfaces relative to an observer. Each row presents stereo pairs demonstrating the induction of the same three-dimensional illusory surface in a different position or orientation, and a schematic view of the percept obtained on stereo-pair fusion (stereopair viewing and schematic conventions as in Fig. 1). *a*, One view of an illusory parabolic surface. *b*, Same surface as in *a*, but positioned further from observer. *c*, Same surface as in *b*, but rotated by 25° . Illusory contours and surfaces are perceived to have the same three-dimensional shape in each case, demonstrating their stability to changes in viewpoint, and our ability to position and orient them in three-dimensional space.

METHODS. Five observers (three naive) were shown coded, randomized stimuli designed to induce the four elementary shapes in at least three different positions and orientations using motion, stereo or both cues to depth. Each stimulus was shown two or more times as a control for successive presentation. Observers were asked to describe their perceptions as completely as possible. Responses were recorded, transcribed, and scored for cue, position, orientation and shape by a naive assistant. Illusory percepts were scored as being 'the same' when an observer reported no difference in position, orientation and shape of successively presented illusory forms. All observers spontaneously noted the recurrence of the same four elementary shapes, and reported seeing the same shapes over changes in cue, position and orientation.

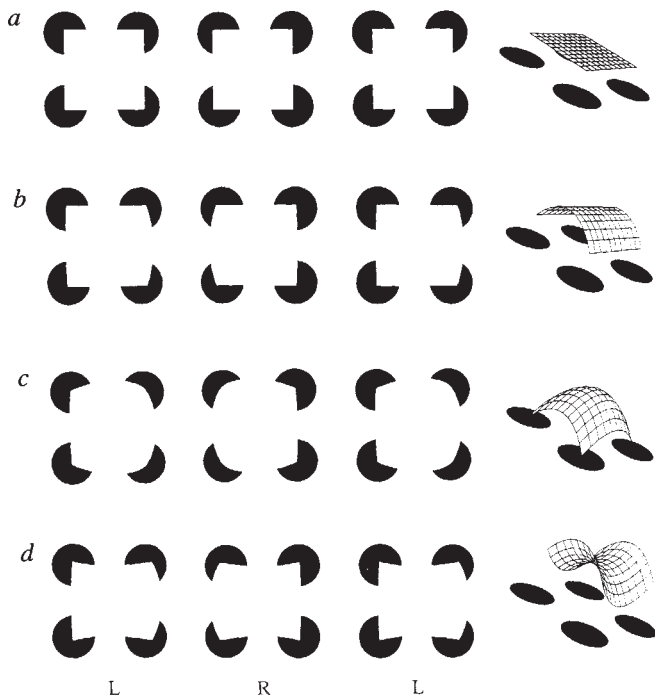


FIG. 3 Morphic generality of three-dimensional illusory contours and surfaces. Each row presents stereo pairs demonstrating the induction of one of the elementary surface shapes, and a schematic view illustrating the percept obtained on stereo fusion (stereo-pair viewing and schematic conventions as in Fig. 1). All smooth, continuous surfaces can be decomposed into regions having one of these four elementary surface shapes, which can be characterized by two orthogonal extremes of curvature, known as the principle curvatures C_1 and C_2 . Their product $C_1 \times C_2$ gives the intrinsic curvature of the surface¹¹. By altering the sign and magnitude of the principle curvatures of an 'invisible' surface which partially occludes four coplanar discs, but which has no other contrast with the surround, we can generate illusory versions of all four elementary surface shapes. *a*, Planar surface with $C_1 = C_2 = 0$. *b*, Parabolic surface with $C_1 = 0$ and $C_2 \neq 0$. *c*, Elliptic surface with $C_1 \times C_2 > 0$. *d*, Hyperbolic surface with $C_1 \times C_2 < 0$. Observed illusory surfaces have shapes which cannot be consistently represented by either minimal surfaces or membranes, which have $C_1 + C_2 = 0$, or thin-plate splines, which have $C_1 \times C_2 = 0$, and therefore cannot result from interpolative mechanisms which make such assumptions²⁶. Nor can the observed illusory surfaces be explained by inference based on generic surfaces, which are unchanged by small perturbations of shape, and therefore do not include planar or parabolic regions²⁷. Note that the three-dimensional illusory surfaces are bounded by three-dimensional contours which are themselves partially illusory¹⁶, and will in general contain sufficient information to reconstruct the global geometry of the surface on which they lie. This suggests that the geometry of the illusory surface can be determined from the geometry specified on such contours through solution of a boundary-value problem (Dirichlet or Neumann problem)²⁸. Any piecewise-smooth surface, including the elementary surfaces shown here, can be reconstructed in this fashion.

can construct illusory contours and surfaces having arbitrary position, orientation and curvature strongly suggests that morphoptic representations depict three-dimensional attributes of visual objects as the local geometry of these illusory forms exists nowhere else but in these representations. Such representations could be implemented by neurons selective for such geometric attributes as three-dimensional position, orientation and curvature, whose activity would support perception of real as well

as illusory contours and surfaces. Although neurons responsive to two-dimensional illusory contours and figures have been described^{24,25}, neuronal response or selectivity for either real or illusory three-dimensional forms of the kind shown here has yet to be demonstrated. Our ability to induce illusory percepts of such varied three-dimensional forms suggests that neurons selective for such geometric attributes may well exist within the visual system. □

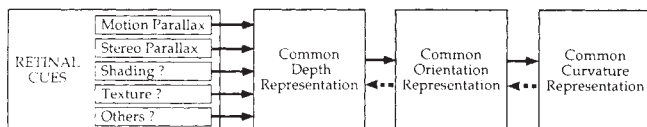


FIG. 4 Proposed mechanism of morphopsis. Various retinal cues, such as motion and stereo parallax, and possible texture and shading⁴, are initially processed separately but then combined into a common, cue-invariant representation of depth^{29,30}. Elements of this representation are explicitly selective for depth relative to the horopter, based primarily on the translational (horizontal and vertical) components of stereo and motion parallax. These elements also carry implicit information about surface orientation, based on the rotational and dilational components of parallax, and about surface shape (curvature), based on higher-order components of parallax. These attributes are made explicit by feedforward transformations (dark arrows) which yield a common representation of orientation in depth (first derivative of depth), and then a common representation of shape or curvature in depth (second derivative of depth) for each point in the visual field. Owing to the geometric interrelationship of these three attributes, each representation constrains the previous one and is constrained by the next (dashed arrows) so that mutually consistent representation of smooth, piecewise-continuous surfaces can be obtained. By representing each of these three attributes as a harmonic function of the visual field, interpolation can be obtained through use of Dirichlet's formula, which contains Green's functions describing the transformations of the visual field necessary to generate each representation²⁸. Although each representation can itself be considered to be a two-dimensional image (scalar function of the visual field), their elements represent attributes that are intrinsically three-dimensional, and constitute a geometric reconstruction of the visual scene. The three representations together are sufficient to account for our ability to induce illusory contours and surfaces having a broad range of three-dimensional positions, orientations and all possible elementary shapes.

Received 9 September; accepted 23 October 1992.

- Schumann, F. E. *Psychol. Studien*, **1**, 1–32 (1904).
- Ehrenstein, W. Z. *Psychol.*, **150**, 83–91 (1941).
- Kanizsa, G. *Riv. Psicol.*, **49**, 7–30 (1955).
- Coren, S. *Psychol. Rev.*, **79**, 359–367 (1972).
- Gregory, R. L. *Nature*, **238**, 51–52 (1972).
- Kanizsa, G. *Scient. Am.*, **234**, 48–52 (1976).
- Kanizsa, G. *Organization in Vision: Essays on Gestalt Perception* (Praeger, New York, 1979).
- Albright, T. D. *Science*, **255**, 1141–1143 (1992).
- DeYoe, T. D. & Van Essen, D. C. *Trends Neurosci.*, **11**, 219–226 (1988).
- Koenderink, J. J. & van Doorn, A. J. *Biol. Cybern.*, **24**, 51–59 (1976).
- Hilbert, D. & Cohn-Vossen, S. *Geometry and the Imagination* (Chelsea, New York, 1952).
- Lawson, R. B. & Gulick, W. L. *Vision Res.*, **7**, 271–297 (1967).
- Klymenko, V. & Weisstein, N. *Perception*, **10**, 627–636 (1981).
- Kellman, P. J. & Cohen, M. H. *Percept. Psychophys.*, **35**, 237–244 (1984).
- Prazdny, K. *Percept. Psychophys.*, **39**, 175–178 (1986).
- Harris, J. P. & Gregory, R. L. *Perception*, **2**, 235–247 (1973).
- Mustillo, P. & Fox, R. *Percept. Psychophys.*, **40**, 362–363 (1986).
- Nakayama, K., Shimojo, S. & Silverman, G. H. *Perception*, **18**, 55–68 (1989).
- Gibson, J. J. *Am. J. Psychol.*, **63**, 376–384 (1950).
- Kellman, P. J. & Shipley, T. F. *Cognitive Psychol.*, **23**, 141–221 (1991).
- Collett, T. S. *Proc. R. Soc. B224*, 43–56 (1985).
- Mitchison, G. J. & McKee, S. *Nature*, **315**, 402–404 (1985).
- Saidpour, A., Braunstein, M. L. & Hoffman, D. D. *Percept. Psychophys.*, **51**, 105–117 (1992).
- von der Heydt, R., Peterhans, E. & Baumgartner, G. *Science*, **224**, 1260–1262 (1984).
- Grosz, D. M., Shapley, R. M. & Hawkins, M. J. *Invest. Ophthalmol. Vis. Sci.*, **33**, 1257 (1992).
- Koch, C., Marroquin, A., & Yuille, A. *Proc. natn. Acad. Sci. U.S.A.*, **83**, 4263–4267 (1986).
- Richards, W. A., Koenderink, J. J. & Hoffman, D. D. *J. opt. Soc. Am.*, **A4**, 1168–1175 (1987).
- Carman, G. J. *Mappings of the Cerebral Cortex* 48–66 (Caltech, Pasadena, 1990).
- Rodgers, B. & Graham, M. *Vision Res.*, **22**, 261–270 (1982).
- Carman, G. J., Welch, L. & Albright, T. D. *Invest. Ophthalmol. Vis. Sci.*, **33**, 1333 (1992).
- Ideasawa, M. *Jap. J. appl. Phys.*, **30**, L751–L754 (1991).

ACKNOWLEDGEMENTS. We thank T. D. Albright, B. Frost, D. Lawrence and T. Sejnowski for contributions and support of this research and T. D. Albright, J. Gallant, D. D. Hoffman, D. I. Macleod, L. Prather, T. Sejnowski, G. R. Stoner and D. C. Van Essen for comments on drafts of this paper. G.J.C. and L.W. were supported by individual National Research Service Awards from the National Eye Institute. G.J.C. was also supported by a Fellowship from the McDonnell-Pew Center for Cognitive Neuroscience, San Diego. Additional support was provided by a grant from the National Eye Institute to T. D. Albright.

Association of dystrophin-related protein with dystrophin-associated proteins in *mdx* mouse muscle

Kiichiro Matsumura, James M. Ervasti, Kay Ohlendieck, Steven D. Kahl & Kevin P. Campbell

Howard Hughes Medical Institute and Department of Physiology and Biophysics, University of Iowa College of Medicine, Iowa City, Iowa 52242, USA

DYSTROPHIN is associated with a complex of muscle membrane (sarcolemmal) glycoproteins that provide a linkage to the extracellular matrix protein, laminin¹⁻⁸. The absence of dystrophin leads to a dramatic reduction of the dystrophin-associated proteins (156DAG, 59DAP, 50DAG, 43DAG and 35DAG) in the sarcolemma of patients with Duchenne muscular dystrophy and *mdx*

mice^{2,6-8}. Here we demonstrate that dystrophin-related protein (DRP, utrophin), an autosomal homologue of dystrophin⁹⁻¹⁷, is associated with an identical or antigenically similar complex of sarcolemmal proteins and that DRP and the dystrophin/DRP-associated proteins colocalize to the neuromuscular junction in Duchenne muscular dystrophy and *mdx* muscle. The DRP and dystrophin/DRP-associated proteins are found throughout the sarcolemma in small-calibre skeletal muscles and cardiac muscle of adult *mdx* mice. Because these muscles show minimal pathological changes¹⁸⁻²⁰, our results could provide a basis for the upregulation of DRP as a potential therapeutic approach.

The immunofluorescence localization of DRP and dystrophin-associated proteins (DAPs) in adult control mouse, adult *mdx* mouse and Duchenne muscular dystrophy (DMD) quadriceps muscle are shown in Fig. 1. In control mouse, dystrophin and DAPs were localized throughout the sarcolemma but were enriched at the neuromuscular junction (NMJ). DRP was localized to the NMJ in control and *mdx* mouse and DMD muscle. As previously reported^{2,6-8}, DAPs were greatly reduced in the

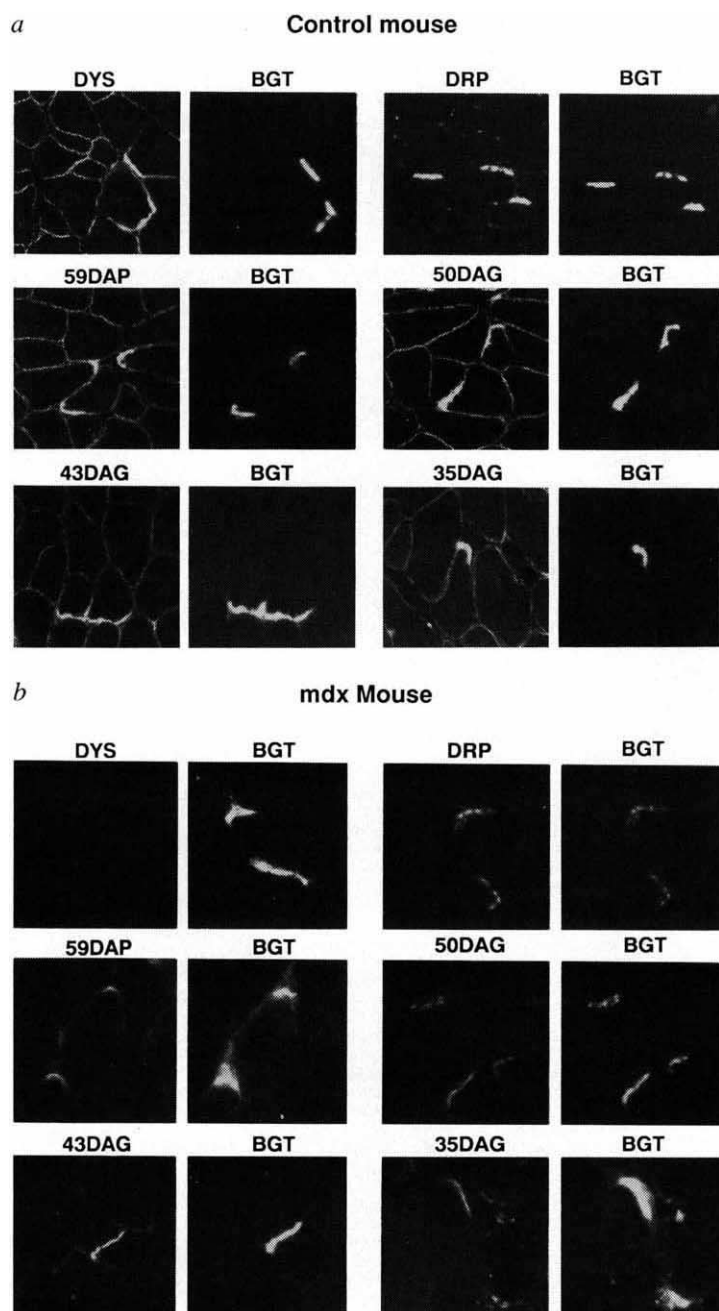
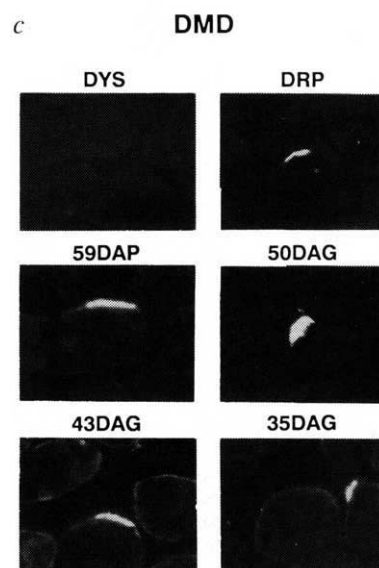


FIG. 1 Immunohistochemical localization of dystrophin (DYS), DRP and DAPs in control mouse (**a**), *mdx* mouse (**b**) and DMD (**c**) skeletal muscle. 156DAG showed the same distribution as the other DAPs (not shown).

METHODS. Cryosections (7 μ m) of quadriceps muscle of adult control mouse, adult *mdx* mouse and DMD were stained with an antibody against dystrophin, DRP or DAPs which was detected with FITC-conjugated secondary antibody^{2,4-8,14}. NMJs were identified by double-staining with rhodamine-labelled α -bungarotoxin (BGT) in control and *mdx* muscles. NMJs in DMD muscle were identified by double staining with BGT or the staining of the serial sections by the Koelle's method.



extrajunctional sarcolemma of *mdx* mouse and DMD muscle. However, near-normal intensity of DAP staining was observed at the NMJ in *mdx* mouse and DMD muscle (Fig. 1b, c).

The colocalization of DRP and DAPs at the NMJ in mouse and DMD muscle suggested a possible association of DRP with DAPs. To test if DRP exists in a complex similar to the dystrophin-glycoprotein complex, isolation of a DRP complex was attempted from both control and *mdx* skeletal muscle membranes using the methods described previously for the isolation of dystrophin-glycoprotein complex: wheat-germ

agglutinin-Sepharose column chromatography of digitonin-solubilized skeletal muscle microsomes followed by DEAE-cellulose column chromatography^{1,2}. DEAE-cellulose eluates were then loaded on 5 to 20% sucrose density gradients for sedimentation analysis (Fig. 2). Because of the much lower relative abundance of DAPs in the *mdx* muscle, three times more *mdx* material was loaded on the sucrose gradient than for the control mouse gradient. Even with the increased sample loading, the immunoblot analysis of the *mdx* sucrose gradient profile required a much longer development time in order to

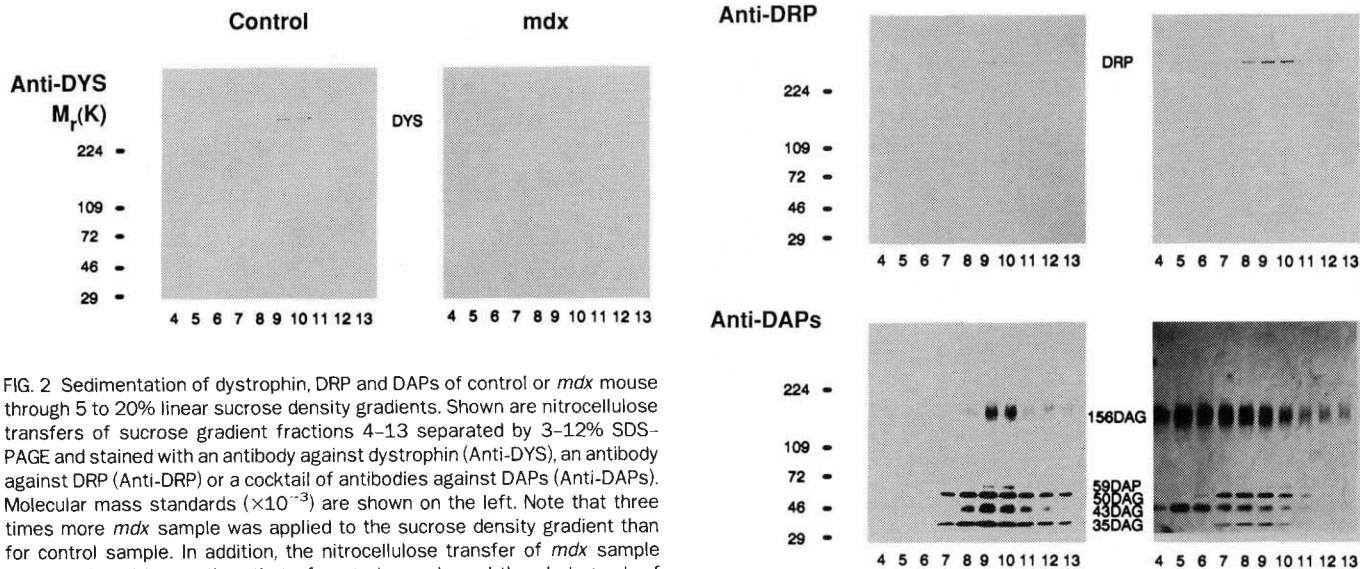
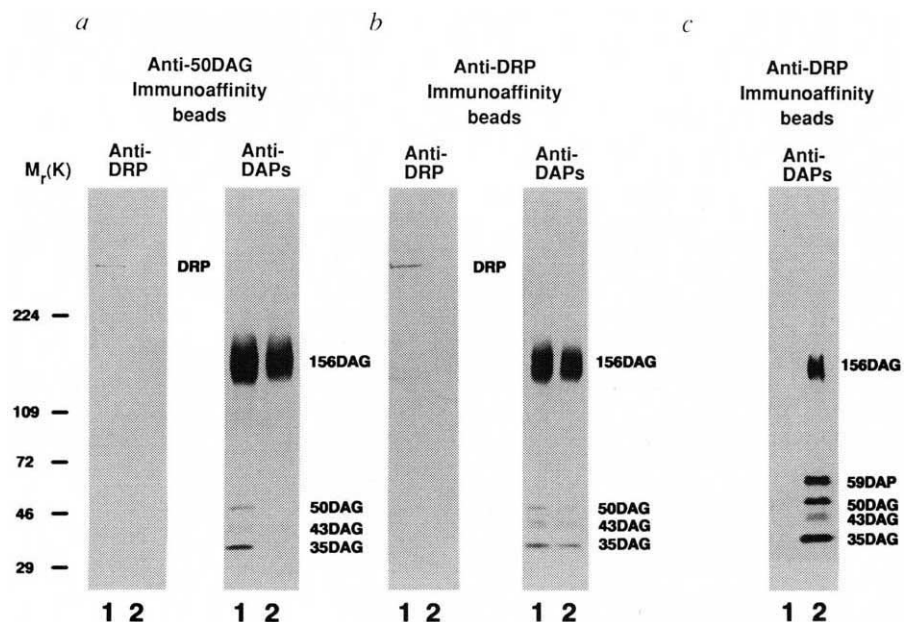


FIG. 2 Sedimentation of dystrophin, DRP and DAPs of control or *mdx* mouse through 5 to 20% linear sucrose density gradients. Shown are nitrocellulose transfers of sucrose gradient fractions 4–13 separated by 3–12% SDS-PAGE and stained with an antibody against dystrophin (Anti-DYS), an antibody against DRP (Anti-DRP) or a cocktail of antibodies against DAPs (Anti-DAPs). Molecular mass standards ($\times 10^{-3}$) are shown on the left. Note that three times more *mdx* sample was applied to the sucrose density gradient than for control sample. In addition, the nitrocellulose transfer of *mdx* sample was developed longer than that of control sample and the photograph of *mdx* sample was overexposed.

METHODS. KCl-washed heavy microsomes (300 mg) prepared from limb and back muscles of adult control or *mdx* mice were solubilized with 280 ml of a solution containing 1% digitonin and 0.5 M NaCl¹⁻³. The solubilized proteins were circulated overnight on a 20 ml wheat-germ agglutinin-Sepharose column. After washing, the column was eluted with 60 ml of a solution containing 0.3 M *N*-acetylglucosamine. The eluate was applied to a 1.5 ml DEAE-cellulose column and eluted with increasing concentrations of NaCl¹⁻³.

FIG. 3 *a*, Immunoabsorption of DRP and DAPs by anti-50DAG immunoaffinity beads. Anti-50DAG immunoaffinity column (goat anti-mouse IgG-Sepharose coupled with anti-50DAG antibody) void (lane 2) or control column (goat anti-mouse IgG-Sepharose) void (lane 1) immunostained with an antibody against DRP (Anti-DRP) or a cocktail of antibodies against DAPs (Anti-DAPs). *b*, Immunoabsorption of DRP and DAPs by anti-DRP immunoaffinity beads. Anti-DRP immunoaffinity column (protein A-Sepharose coupled with anti-DRP antibody) void (lane 2) or control column (protein A-Sepharose) void (lane 1) immunostained with an antibody against DRP or a cocktail of antibodies against DAPs. The 59DAP (which is not glycosylated) is not identified clearly in (*a*) and (*b*) owing to its extremely low abundance in the wheat-germ agglutinin (WGA) eluate. *c*, Immunoabsorption of DAPs by anti-DRP immunoaffinity beads. Proteins adsorbed by anti-DRP immunoaffinity column (lane 2) or control column (lane 1) immunostained with a cocktail of antibodies against DAPs. 59DAP is clearly detected because it was concentrated on the immunoaffinity beads saturated with the WGA eluate. Molecular mass standards ($\times 10^{-3}$) are shown on the left.

METHODS. For *a* and *b*, 100 μ l of 0.3 M *N*-acetylglucosamine eluate from WGA-Sepharose column chromatography of *mdx* mouse (0.2 mg ml⁻¹) was incubated with 30 μ l of either immunoaffinity or control beads² in the presence of 0.5 M NaCl overnight at 4 °C. After centrifugation, proteins



remaining in the supernatants were analysed by 3–12% SDS-PAGE and immunoblotting. For *c*, the beads (100 μ l) were incubated with a saturating volume (3 ml) of the same material. After centrifugation and extensive washing, proteins adsorbed by the beads were analysed by 3–12% SDS-PAGE and immunoblotting.

detect DAPs. The control mouse sucrose gradient profile demonstrated that dystrophin, DRP and DAPs cosedimented with a peak in fractions 9 and 10. The *mdx* mouse sucrose gradient profile revealed that DRP and 59DAP also cosedimented with a peak in fractions 9 and 10. In contrast, 156DAG and 43DAG cosedimented in a broad range of fractions, including fractions 9 and 10, whereas 50DAG and 35DAG cosedimented with a peak in fractions 8 and 9. These results suggest that there are two populations of DAPs in *mdx* muscle: a small fraction of DAPs associated with DRP to form a complex similar in size to dystrophin-glycoprotein complex and a large fraction of free DAPs which fail to form a complex in the absence of dystrophin and therefore sediment as much smaller entities.

To test the possible association of DRP with DAPs more directly, immunoadsorption experiments of wheat-germ agglutinin eluate from solubilized *mdx* muscle membranes were done using immunoaffinity beads prepared with antibodies against 50DAG or DRP (Fig. 3). Proteins remaining in the supernatants after centrifugation of the beads were analysed by SDS-polyacrylamide gel electrophoresis and immunoblotting. Coomassie blue-staining of the gel revealed that the overall protein composition in the supernatants was indistinguishable

between immunoaffinity and control beads in both anti-50DAG and anti-DRP immunoprecipitation (not shown). Immunoblot analysis revealed that anti-50DAG beads precipitated all of the 50DAG, 35DAG and DRP as well as a small fraction of 156DAG and 43DAG (Fig. 3a). Conversely, anti-DRP beads precipitated all DRP and a small fraction of DAPs (Fig. 3b). Autoradiography of the immunoblots using 125 I-labelled protein A revealed that about 30% of DAPs were precipitated by the anti-DRP beads except 59DAP, which was completely precipitated. Immunoblot analysis of anti-DRP beads incubated with a volume of wheat-germ agglutinin eluate sufficient to saturate all DRP antibody-binding sites confirmed that all of the DAPs could be precipitated by anti-DRP beads (Fig. 3c). Similar results were obtained using the wheat-germ agglutinin eluate from solubilized control mouse muscle membranes: (1) anti-50DAG beads precipitated both dystrophin and DRP together with all DAPs and (2) anti-DRP beads precipitated DRP together with a small fraction of DAPs (not shown).

Thus far, our results indicate that there are two populations of DAPs in *mdx* muscle: a small fraction associated with DRP at the NMJ and an uncomplexed fraction which, we believe, would normally associate with dystrophin. The components of

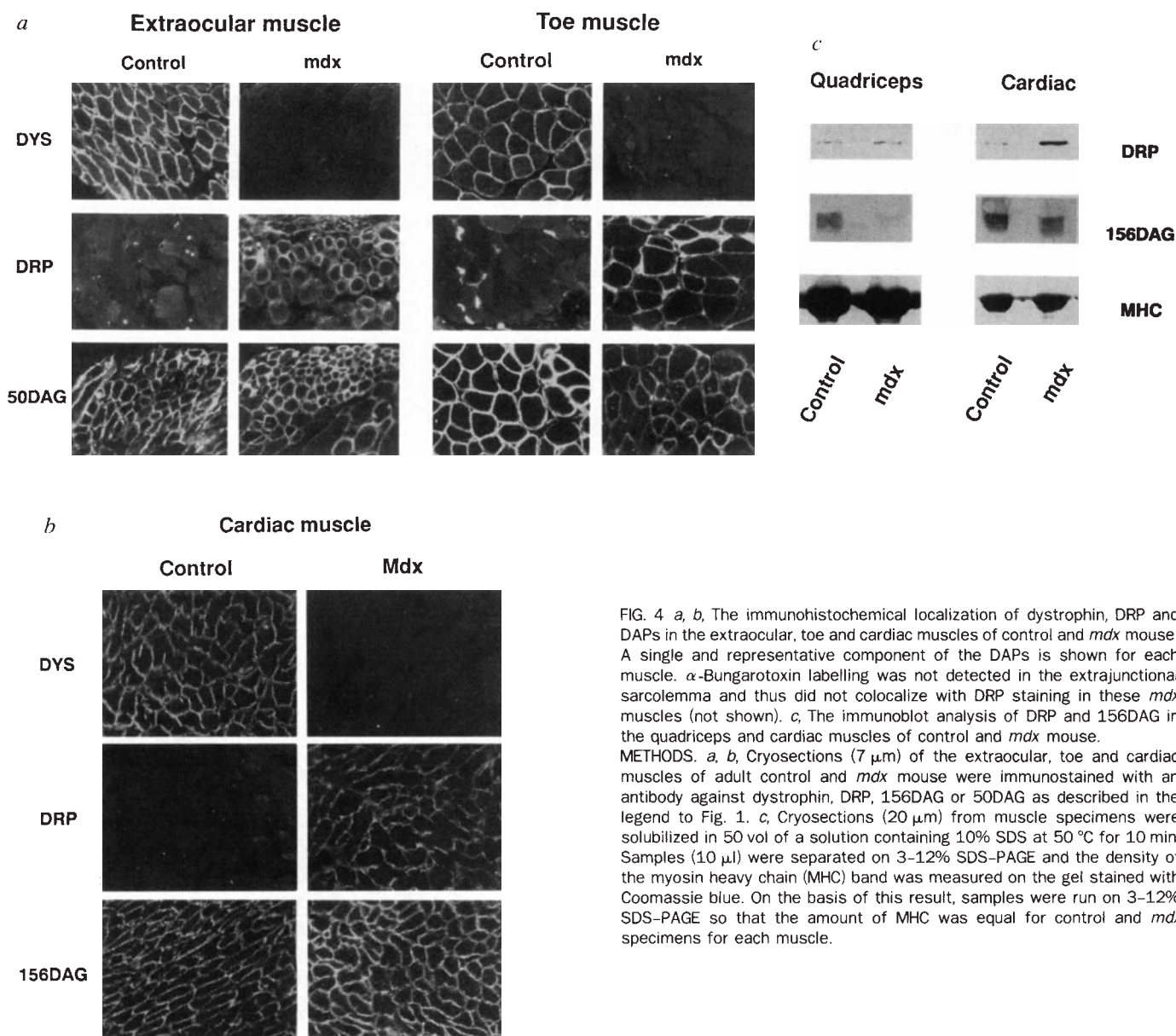


FIG. 4 a, b, The immunohistochemical localization of dystrophin, DRP and DAPs in the extraocular, toe and cardiac muscles of control and *mdx* mouse. A single and representative component of the DAPs is shown for each muscle. α -Bungarotoxin labelling was not detected in the extrajunctional sarcolemma and thus did not colocalize with DRP staining in these *mdx* muscles (not shown). c, The immunoblot analysis of DRP and 156DAG in the quadriceps and cardiac muscles of control and *mdx* mouse. METHODS. a, b, Cryosections (7 μ m) of the extraocular, toe and cardiac muscles of adult control and *mdx* mouse were immunostained with an antibody against dystrophin, DRP, 156DAG or 50DAG as described in the legend to Fig. 1. c, Cryosections (20 μ m) from muscle specimens were solubilized in 50 vol of a solution containing 10% SDS at 50 $^{\circ}$ C for 10 min. Samples (10 μ l) were separated on 3–12% SDS-PAGE and the density of the myosin heavy chain (MHC) band was measured on the gel stained with Coomassie blue. On the basis of this result, samples were run on 3–12% SDS-PAGE so that the amount of MHC was equal for control and *mdx* specimens for each muscle.

these two populations could be identical gene products or antigenically cross-reactive isoforms or homologues.

It is well known that all muscles in DMD or in *mdx* mouse are not affected to the same degree although dystrophin is deficient in all muscles¹⁸⁻²¹. For instance, dysfunction of small-calibre skeletal muscles is minimal in both DMD and *mdx* mouse^{18,19} and dysfunction of cardiac muscle is absent in *mdx* mouse²⁰. The association of DRP with DAPs, together with the recent finding that DRP has been detected in the sarcolemma outside the NMJ in DMD or *mdx* mouse¹⁵⁻¹⁷, suggest the possibility that DRP could compensate for dystrophin deficiency by retaining DAPs in extrajunctional regions of the sarcolemma. To investigate this possibility, we did immunohistochemical analysis of small skeletal muscles and cardiac muscle of adult *mdx* mouse (Fig. 4a, b). DRP and DAPs are expressed throughout the sarcolemma (compare Fig. 1b), whereas in the same muscles in control mouse only dystrophin and DAPs are expressed throughout the sarcolemma. In addition, immunoblot analysis of the crude muscle extracts and its quantitation by autoradiography show a fourfold increase of DRP and a near-normal level of 156DAG in *mdx* cardiac muscle, whereas there is a 1.3-fold increase of DRP and a drastic reduction of 156DAG in *mdx* quadriceps muscle compared to the respective control muscles (Fig. 4c). The preservation of DAPs in small skeletal

and cardiac muscles of *mdx* mouse suggests that altered expression of DRP could lead to the retention of DAPs. □

Received 17 August; accepted 6 October 1992.

- Campbell, K. P. & Kahl, S. D. *Nature* **338**, 259-262 (1989).
- Ervasti, J. M., Ohlndieck, K., Kahl, S. D., Gaver, M. G. & Campbell, K. P. *Nature* **345**, 315-319 (1990).
- Ervasti, J. M., Kahl, S. D. & Campbell, K. P. *J. Biol. Chem.* **266**, 9161-9165 (1991).
- Ohlndieck, K., Ervasti, J. M., Snook, J. B. & Campbell, K. P. *J. Cell Biol.* **112**, 135-148 (1991).
- Ervasti, J. M. & Campbell, K. P. *Cell* **66**, 1121-1131 (1991).
- Ohlndieck, K. & Campbell, K. P. *J. Cell Biol.* **115**, 1685-1694 (1991).
- Ibraghimov-Beskrovnaya, O. et al. *Nature* **355**, 696-702 (1992).
- Ohlndieck, K. et al. *Neurology* (in press).
- Love, D. R. et al. *Nature* **339**, 55-58 (1989).
- Buckle, V. J., Guenet, J. L., Simon-Chazottes, D., Love, D. R. & Davies, K. E. *Hum. Genet.* **85**, 324-326 (1990).
- Khurana, T. S., Hoffman, E. P. & Kunkel, L. M. *J. Biol. Chem.* **265**, 16717-16720 (1990).
- Love, D. R. et al. *Proc. natn. Acad. Sci. U.S.A.* **88**, 3243-3247 (1991).
- Pons, F. et al. *FEBS Lett.* **282**, 161-165 (1991).
- Ohlndieck, K. et al. *Neuron* **7**, 499-508 (1991).
- Khurana, T. S. et al. *Neuromusc. Dis.* **1**, 185-194 (1991).
- Man, N. et al. *J. Cell Biol.* **115**, 1695-1700 (1991).
- Takemitsu, M. et al. *Biochem. biophys. Res. Commun.* **180**, 1179-1186 (1991).
- Karpati, G. & Carpenter, S. *Am. J. med. Genet.* **25**, 653-658 (1986).
- Karpati, G., Carpenter, S. & Prescott, S. *Muscle Nerve* **11**, 795-803 (1988).
- Torres, L. F. B. & Duchon, L. W. *Brain* **110**, 269-299 (1987).
- Matsumura, K. et al. *Muscle Nerve* **11**, 97-102 (1988).

ACKNOWLEDGEMENTS. We thank S. W. Sernett and C. J. Leveille for technical assistance and T. Partridge (Westminster Medical School) for comments. K.P.C. is an Investigator of the Howard Hughes Medical Institute. J.M.E. is the Carl M. Pearson Fellow of the Muscular Dystrophy Association. This work was also supported by the Muscular Dystrophy Association.

Primary structure of dystrophin-related protein

J. M. Tinsley, D. J. Blake, A. Roche, U. Fairbrother*, J. Riss, B. C. Byth†, A. E. Knight‡, J. Kendrick-Jones‡, G. K. Suthers, D. R. Love†, Y. H. Edwards* & K. E. Davies§

Institute of Molecular Medicine, John Radcliffe Hospital, Headington, Oxford OX3 9DU, UK

* Galton Laboratory, MRC Human Biochemical Genetics Laboratory, Wolfson House, 4 Stephenson Way, London NW1 2HE, UK

‡ MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

DYSTROPHIN-RELATED protein (DRP or 'utrophin')¹ is localized in normal adult muscle primarily at the neuromuscular junction²⁻⁴. In the absence of dystrophin in Duchenne muscular dystrophy (DMD) patients, DRP is also present in the sarcolemma³⁻⁷. DRP is expressed in fetal and regenerating muscle and may play a similar role to dystrophin in early development^{3,7-9}, although it remains to be determined whether DRP can functionally replace dystrophin in adult tissue. Previously we described a 3.5-kilobase complementary DNA clone that exhibits 80 per cent homology to the C-terminal domain of dystrophin¹⁰. This sequence identifies a 13-kilobase transcript that maps to human chromosome 6 (refs 2, 11). Antibodies raised against the gene product identify a polypeptide with a relative molecular mass of about 400K in all tissues examined^{7,8,12}. To investigate the relationship between DRP and dystrophin in more detail, we have cloned and sequenced the whole DRP cDNA. Homology between DRP and dystrophin extends over their entire length, suggesting that they derive from a common ancestral gene. Comparative analysis of primary sequences highlights regions of functional importance, including those that may mediate the localization of DRP and dystrophin in the muscle cell.

Because DRP is under-represented in amplified cDNA libraries, we constructed an unamplified library using messenger

RNA extracted from the human glioma cell line IN157, in which DRP is abundantly expressed in the absence of dystrophin. Initially the library was screened with a 1.9-kb fragment derived from the B3 cDNA referred to previously, thereafter a stepwise screening procedure led to the isolation of nine overlapping clones covering the complete DRP cDNA. A single large open reading frame of 10,299 base pairs (bp) is preceded by an ATG initiation codon and 550 bp of 5' untranslated sequence. The stop codon is followed by about 2 kb of 3' untranslated region containing an AATAAA poly(A) consensus sequence. The open reading frame encodes a protein of 3,433 amino acids with a predicted relative molecular mass of 395K a size similar to dystrophin *M_r* 427K).

DNA and amino-acid sequence comparisons of DRP and dystrophin show the overall structure to be similar with a putative actin-binding domain in the first 250 amino acids, a long region containing multiple repeats followed by a cysteine-rich domain¹³. In the first 250 amino acids, 80 per cent of the residues are conserved between DRP and dystrophin. This region of dystrophin is known to bind actin *in vitro*¹⁴⁻¹⁶ and computer analysis identified two actin-type actin-binding motifs at amino-acid positions 33-42 and 107-131 in DRP. Figure 1 shows an amino-acid comparison between DRP, dystrophin and the consensus sequence for α -actinin¹⁴. The proposed 15-amino-acid actin-binding domain of α -actinin¹⁷ is also conserved in dystrophin and DRP with nine exact matches and three conserved matches. These data suggest that DRP, like dystrophin, binds actin and consequently may have a similar cellular function to that of dystrophin.

DRP and dystrophin are similar throughout most of their lengths, although DRP lacks two large regions in the second half of the rod domain. Excessive emphasis has been placed on the 'spectrin repeats' of dystrophin as the repeats are much more variable than those of spectrin, which suggests that their structure may differ. The predictions of coiled-coil regions are very closely related between human, mouse and chicken dystrophin sequences (data not shown). Although a broadly similar pattern is seen in DRP, there are significant differences in the rod domain. Whether these predictions are consistent with existing models for the structure of dystrophin is not clear, but they suggest that the structure is more irregular than previously described. This may reflect lower evolutionary constraints on the structure of this region in the two proteins. Indeed, large

† Present addresses: CRC Human Cancer Genetics Research Group, Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QP, UK (D.R.L.); The Hospital for Sick Children, 555 University Avenue, Toronto, Ontario, Canada (B.C.B.).

§ To whom correspondence should be addressed.

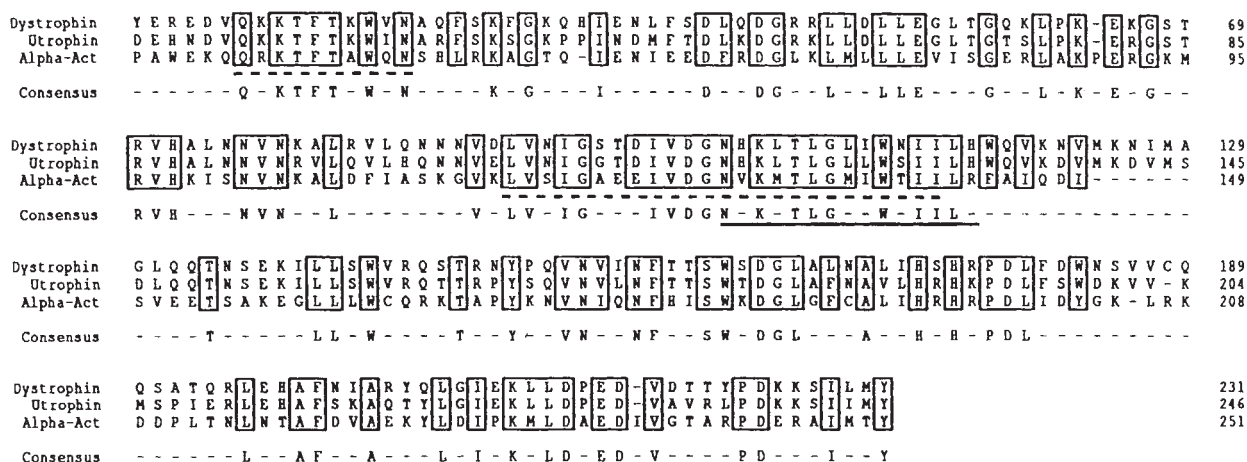


FIG. 1 Sequence comparison between the N terminus of human dystrophin¹³ (residues 11–231), DRP (utrophin; residues 27–246) and the α -actinin consensus sequence¹⁴. The sequences were aligned with the aid of the computer program PILEUP (UWGCG²³); boxes enclose conserved residues

and the consensus line indicates regions of exact homology. Bold broken lines indicate regions important in actin binding (MOTIFS, UWGCG) and the region underlined is the 15-amino-acid actin-binding domain of α -actinin¹⁷.

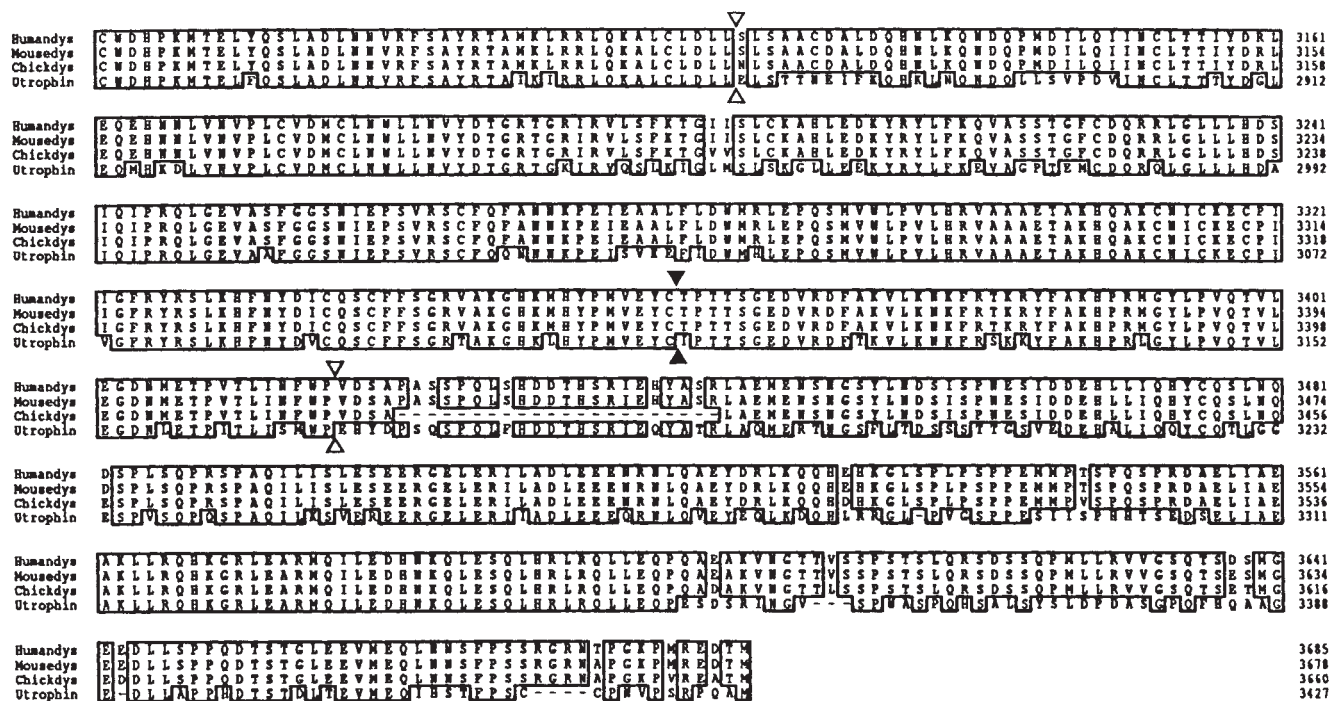


FIG. 2 Amino-acid sequence comparison of the C termini of human dystrophin¹³ (3,075–3,685), mouse dystrophin (3,075–3,678; J. S. Chamberlain, J. Pearlman, D. Muzny and N. Farwell, unpublished data), chicken dystrophin²⁴ (3,079–3,660) and DRP (utrophin; 2,833–3,427). Sequences were aligned using the program PILEUP (UWGCG). Blocked regions indicate

exact conservation of residues. Solid arrowheads mark the boundary between the cysteine-rich and C-terminal domains. Two open arrowheads delimit the putative binding region with the dystrophin-associated glycoproteins²⁰.

deletions of this domain in dystrophin result in only a very mild phenotype¹⁸. An interesting feature at the very end of the carboxy terminus is the pair of predicted coiled-coil regions which are virtually identical in all four published sequences. The function of this region is unknown.

In dystrophin, part of the C-terminal domain and the adjacent cysteine-rich region are involved in the binding of the dystrophin glycoprotein complex^{19,20}. Absence of the carboxy terminus of dystrophin is associated with severe phenotypes in most muscular dystrophy patients^{21,22}. Comparison of the C-terminal

regions of DRP and the dystrophins (Fig. 2) shows that this region is highly conserved with an 80 per cent exact match and about half of the remaining residues conserved. Strong conservation of the glycoprotein-binding region suggests that DRP may bind to the sarcolemma in dystrophic muscle through the glycoprotein complex in a similar manner to dystrophin. It remains to be seen whether interaction occurs between DRP and glycoproteins in tissues other than muscle and whether such interaction is important to the function of DRP in these tissues. □

Received 22 September; accepted 29 October 1992.

1. Blake, D. J. et al. *Hum. molec. Genet.* **1**, 103-109 (1992).
2. Love, D. R. et al. *Proc. natn. Acad. Sci. U.S.A.* **88**, 3243-3247 (1991).
3. Khurana, T. S. et al. *Neuromuscular Dis.* **1**, 185-194 (1991).
4. Pons, F. et al. *FEBS Lett.* **282**, 161-165 (1991).
5. Ohlndiek, K. et al. *Neuron* **7**, 499-508 (1991).
6. Tanaka, H. et al. *Histochemistry* **96**, 1-5 (1991).
7. thilman N. et al. *J. Cell Biol.* **115**, 1695-1700 (1991).
8. Voit, T., Haas, K., Leger, J. O., Pons, F. & Leger, J. J. *Am. J. Path.* **139**, 969-976 (1991).
9. Takemitsu, M. et al. *Biochem. Biophys. Res. Commun.* **180**, 1179-1186 (1991).
10. Love, D. R. et al. *Nature* **339**, 55-58 (1989).
11. Khurana, T. S., Hoffman, E. P. & Kunkel, L. M. *J. biol. Chem.* **265**, 16717-16720 (1990).
12. Augier, N. et al. *J. Neurol. Sci.* **107**, 233-238 (1992).
13. Koenig, M., Monaco, A. P. & Kunkel, L. M. *Cell* **53**, 219-228 (1988).
14. Hemmings, L., Kuhlman, P. A. & Critchley, D. R. *J. Cell Biol.* **116**, 1369-1380 (1992).

15. Way, M. et al. *FEBS Lett.* **301**, 243-245 (1992).
16. Levine, B. A., Moir, A. J., Patchell, V. B. & Perry, S. V. *FEBS Lett.* **263**, 159-162 (1990).
17. Kuhlman, P. A., Hemmings, L. & Critchley, D. R. *FEBS Lett.* **304**, 201-206 (1992).
18. England, S. B. et al. *Nature* **342**, 180-182 (1990).
19. Ervasti, J. M. & Campbell, K. P. *Cell* **66**, 1121-1131 (1991).
20. Suzuki, A., Yoshida, M., Yamamoto, H. & Ozawa, E. *FEBS Lett.* **308**, 154-160 (1992).
21. Arahata, K. et al. *J. neurol. Sci.* **101**, 148-156 (1991).
22. Koenig, M. et al. *Am. J. hum. Genet.* **45**, 498-506 (1989).
23. Devereux, J., Haekerli, P. & Smithies, O. *Nucleic Acids Res.* **12**, 387-395 (1984).
24. Lemaire, C., Heilig, R. & Mandel, J. L. *Nucleic Acids Res.* **16**, 11815 (1988).

ACKNOWLEDGEMENTS. J.M.T. and D.J.B. contributed equally to this work. This work was supported by the Muscular Dystrophy Group of Great Britain, Medical Research Council of Great Britain and the Muscular Dystrophy Association USA. The nucleotide sequence of DRP (utrophin) has been submitted to EMBL under accession number X69086.

CD1b restricts the response of human CD4⁺8⁺ T lymphocytes to a microbial antigen

Steven Porcelli, Craig T. Morita & Michael B. Brenner

Lymphocyte Biology Section, Department of Rheumatology/Immunology, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts 02115, USA

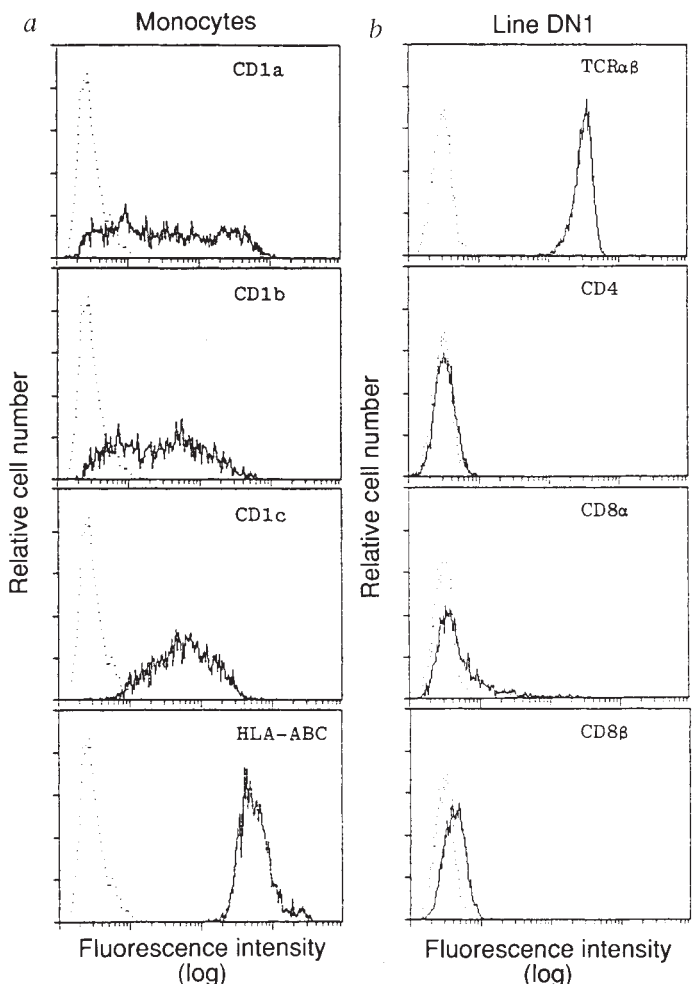
MOLECULES encoded by the human CD1 locus on chromosome 1 (ref. 33) are recognized by selected CD4⁺8⁺ T-cell clones expressing either $\alpha\beta$ or $\gamma\delta$ T-cell antigen receptors^{1,2}. The known structural resemblance of CD1 molecules to antigen-presenting molecules encoded by major histocompatibility complex (MHC) genes on human chromosome 6 (refs 3, 4, 34, 35), suggested that

CD1 may represent a family of antigen-presenting molecules separate from those encoded in the MHC^{1,5,6}. Here we report that the proliferative and cytotoxic responses of human CD4⁺8⁺ $\alpha\beta$ TCR⁺ T cells specific for *Mycobacterium tuberculosis* can be restricted by CD1b, one of the four identified protein products of the CD1 locus. The responses of these T cells to *M. tuberculosis* seemed not to involve MHC encoded molecules, but were absolutely dependent on the expression of CD1b by the antigen-presenting cell and involved an antigen processing requirement similar to that seen in MHC class II-restricted antigen presentation. These results provide, to our knowledge, the first direct evidence for the proposed antigen-presenting function of CD1 molecules and suggest that the CD1 family plays a role in cell-mediated immunity to microbial pathogens.

In studies aimed at developing a system to detect antigen presentation by CD1 molecules, we tested the ability of recombinant cytokines to induce expression of CD1a, -b and -c on peripheral blood monocytes, which normally do not express

FIG. 1 Expression of CD1a, -b and -c by monocytes cultured with GM-CSF and IL-4, and surface phenotype of CD1b-restricted T cells specific for *M. tuberculosis*. **a**, Flow cytometric analysis of peripheral blood monocytes cultured for 60 h in medium containing GM-CSF and IL-4 showing expression of CD1a, -b and -c. Cells were stained with control antibody (dotted line) or antibodies with the specificity indicated in each histogram box (solid lines). Monocytes cultured in the absence of cytokines or with interferon- γ did not express significant levels of CD1a, -b or -c (data not shown). **b**, Flow cytometric analysis of T-cell line DN1 showing expression of $\alpha\beta$ TCRs, absence of CD4, and minimal or absent expression of CD8 (dotted and solid lines represent control and specific antibodies as in **a**).

METHODS. Flow cytometry was done as described²² using the following antibodies: P3 (IgG1 control; ref. 22), OKT6 (anti-CD1a; ref. 23), 4A7.6 (anti-CD1b; ref. 20), 10C3 (anti-CD1c; ref. 24), W6/32 (anti-HLA-ABC; ref. 25), BMA031 (anti- $\alpha\beta$ TCR; ref. 26), OKT4 (anti-CD4; ref. 23), OKT8 (anti-CD8 α ; ref. 23) and 2ST8-5H7 (anti-CD8 β ; ref. 27). Monocytes were isolated from leukocyte concentrates of normal donors by plastic adherence²⁸, and detached by incubation at 37 °C in PBS with 0.53 mM EDTA (PBS/EDTA). Adherent cells were typically >90% CD14⁺ and MHC class II⁺, and negative for CD1a, -b and -c by surface staining (data not shown). To induce CD1 expression, monocytes were cultured for 60 h in RPMI-1640 (Gibco) + 10% fetal calf serum (FCS, Hyclone) with 100 units ml⁻¹ each of GM-CSF and IL-4, and were collected using PBS/EDTA as above. T-cell line DN1 was established from a random normal donor's peripheral blood. Nonadherent mononuclear cells were treated with antibodies OKT4 and OKT8 and rabbit complement, and the remaining viable cells were suspended in a mixture of OKT4, OKT8 and anti-TCR δ 1 (ref. 6) antibodies for 1 h, washed and incubated for 30 min at 4 °C with goat anti-mouse immunoglobulin-coupled magnetic beads (Dynal). After magnetic separation, the resulting CD4⁺8⁺ δ TCR⁺ cells were cultured with equal numbers of autologous monocytes in complete medium (RPMI-1640 with 10% FCS and additional supplements as previously described²⁹) with 100 U ml⁻¹ each of GM-CSF and IL-4. *M. tuberculosis* soluble extract, produced by sonication of desiccated bacilli (strain H37Ra (Difco)) in PBS followed by centrifugation at 100,000 *g* to remove insoluble material, was added to a bacterial protein concentration of 10 μ g ml⁻¹. Cultures were restimulated every 10 to 14 days with *M. tuberculosis* and heterologous CD1⁺ monocytes (CD1 induced as described above) in complete medium, and were fed every three to four days with fresh medium containing 1 nM recombinant interleukin-2 (IL-2).



significant levels of these molecules⁷. High levels of CD1a, -b and -c were consistently observed on monocytes cultured with a combination of granulocyte/monocyte colony stimulating factor (GM-CSF) and interleukin-4 (IL-4) (Fig. 1a). Because monocytes are efficient antigen-presenting cells (APCs), we reasoned that CD1⁺ monocytes might stimulate a CD1-restricted

T-cell response to an exogenous antigen. Given that most CD1-specific T cells identified to date have a CD4⁺8⁻ phenotype^{1,2}, we focused on this subset and generated a T-cell line by repeated stimulation of peripheral blood $\alpha\beta$ TCR⁺ CD4⁺8⁻ T cells with a soluble extract of *M. tuberculosis* and heterologous CD1⁺ monocytes (Fig. 1b). Functional studies of the resulting T-cell

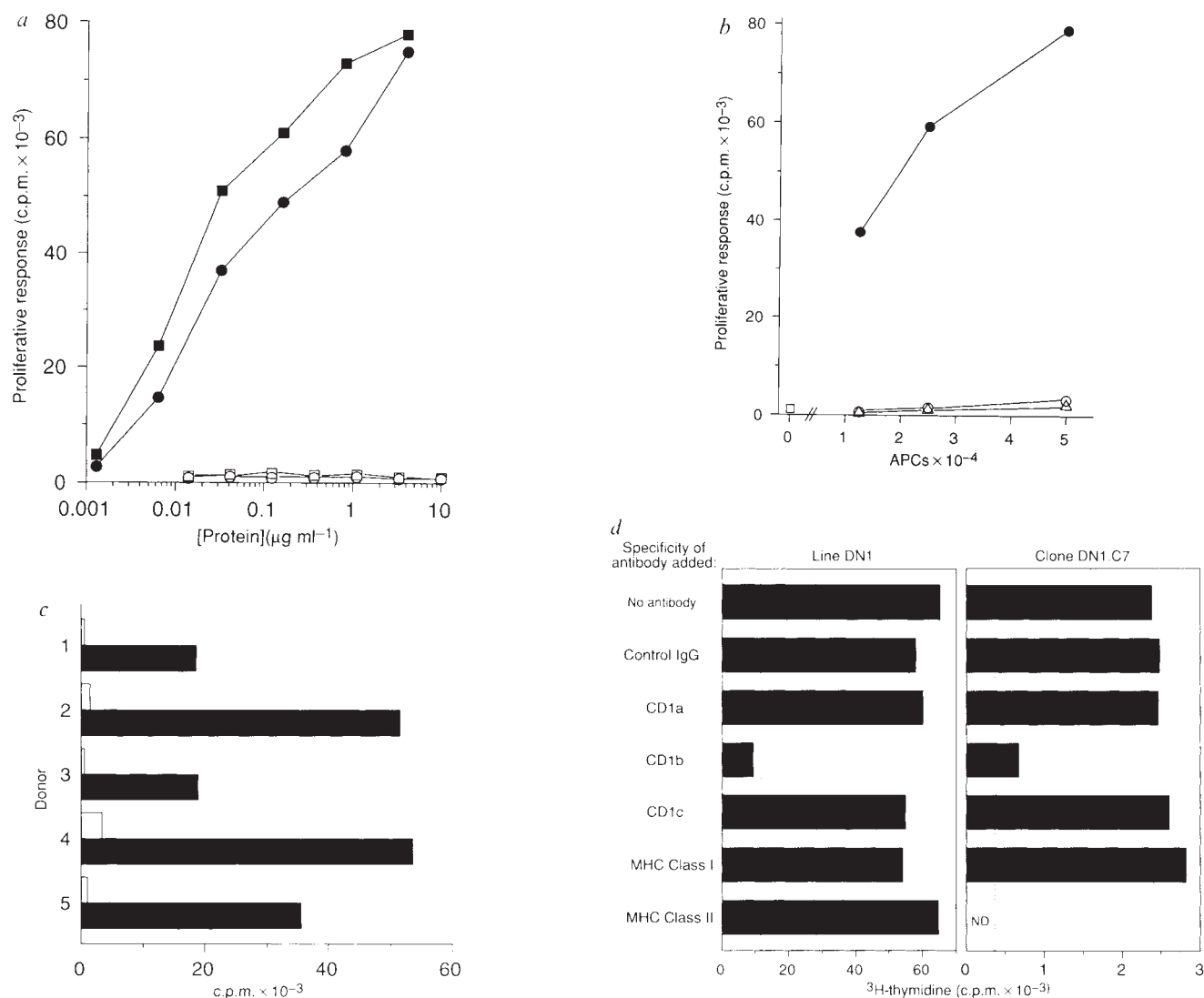


FIG. 2 Antigen specificity and self restriction of proliferative responses of CD4⁺8⁻ T-cell line DN1 and its subclone DN1.C7. **a**, Proliferative responses (counts per minute (c.p.m.) of ³H-thymidine incorporated) of DN1 to *M. tuberculosis* (■), *M. leprae* (●), *Escherichia coli* (○) and tetanus toxoid (□). Antigen-presenting cells were heterologous GM-CSF + IL-4-treated CD1⁺ monocytes. Antigen concentration (based on protein content) is shown on the x-axis. Although the identity of the relevant mycobacterial antigen is not yet established, preliminary experiments showed it to be a nondialysable (that is, molecular mass greater than 12–15K) macromolecule that was precipitated by 50% NH₄SO₄. Purified recombinant mycobacterial 65K and 70K heat shock proteins did not stimulate proliferation of DN1 or DN1.C7 (S.P., unpublished data). **b**, Proliferative response of line DN1 to *M. tuberculosis* (1 μg protein per ml) required CD1⁺ APCs. APCs indicated by symbols as follows: no APCs, □; GM-CSF + IL-4-treated monocytes (CD1⁺), ●; IFNγ-treated monocytes (CD1⁺), ○; freshly isolated monocytes (CD1⁺), △. The number of APCs added to each culture is shown on the x-axis. **c**, APCs from all donors tested supported the proliferative response of line DN1 to *M. tuberculosis*. Open bars, T cells plus APCs without *M. tuberculosis*; solid bars, T cells plus APCs with *M. tuberculosis* (1 μg protein per ml). APCs were GM-CSF + IL-4-treated peripheral blood mononuclear cells from five unrelated donors. HLA typing confirmed that no allele of the HLA-A, -B, -C, -DR, -DP or -DQ loci was shared among all five donors (data not shown). **d**, Anti-CD1b antibody specifically inhibited the proliferative response of DN1

and DN1.C7 to *M. tuberculosis* (1 μg protein per ml). APCs were GM-CSF + IL-4-treated monocytes. Dotted line represents response in the absence of *M. tuberculosis*. Antibodies used were P3 (control IgG), OKT6 (CD1a), WM-25 (CD1b; ref. 30), 10C3 (CD1c), W6/32 (MHC class I), and IVA12 (MHC class II; ref. 31). ND, not determined.

METHODS. Proliferation assays were carried out in triplicate with 5×10^4 each of T cells and irradiated (5,000 Rad) APCs in 200 μl complete medium in 96-well flat-bottom microtitre plates (Linbro). *M. leprae* and *E. coli* soluble extracts were produced as described for *M. tuberculosis*. Monoclonal antibodies were added as purified immunoglobulin to a final concentration of 25 μg ml⁻¹. Cultures were collected on day five (day three for antibody blocking) after a 6 h pulse with 1 μCi ³H-thymidine (6.7 Ci mmol⁻¹, New England Nuclear), and ³H incorporation determined by liquid scintillation counting. Results are expressed as the mean c.p.m. of ³H-thymidine incorporation of triplicate cultures. Isolated monocytes or whole PBMC were treated with recombinant GM-CSF and IL-4 as in legend to Fig. 1, or with IFNγ 100 U ml⁻¹ for 60 h before combining them with T cells for proliferation assay. T-cell clone DN1.C7 was representative of four extensively characterized subclones derived from DN1 by PHA stimulation in limiting dilution culture and propagated using PHA stimulation and IL-2 as described²⁹. All clones derived from line DN1 had a surface phenotype indistinguishable from that shown in Fig. 1b.

line (designated DN1) showed that these T cells gave specific proliferative responses to antigens derived from *M. tuberculosis* and from closely related *M. leprae* bacilli, but not to unrelated bacterial antigens (Fig. 2a). These responses were dependent on monocytes pretreated with GM-CSF+IL-4 (Fig. 2b), and were not restricted by polymorphic MHC determinants (Fig. 2c). This lack of MHC restriction was consistent with restriction by nonpolymorphic CD1 molecules, and we examined this possibility by testing the effects of monoclonal antibodies specific for CD1 or MHC molecules on the *M. tuberculosis*-induced proliferation of line DN1 and a representative subclone, DN1.C7. Only anti-CD1b antibody showed significant blocking, and no consistent effects were observed with anti-CD1a or -c antibodies or with antibodies against monomorphic determinants of MHC class I or class II molecules (Fig. 2d).

The apparent restriction of the *M. tuberculosis*-specific response by CD1b was confirmed using B-cell lines transfected with expression vector constructs encoding CD1 molecules. Analysis of a series of CD1b transfectants of B-cell lines revealed that they were unable to stimulate antigen-specific proliferative responses of line DN1, apparently because of the absence from B-cell lines of an essential membrane-bound costimulatory ligand (S.P., unpublished data). But experiments using B-cell transfectants as targets for cytolytic T-cell activity were informative. Using the B lymphoblastoid cell line C1R (ref. 8), we generated stable transfectants expressing CD1a, -b or -c at

comparable levels and tested each for its ability to present *M. tuberculosis* in a cytolytic assay. Only C1R cells transfected with CD1b and incubated with *M. tuberculosis* before the assay were lysed by line DN1 and its subclone DN1.C7 (Fig. 3a, b). The specificity of this CD1b-restricted response was confirmed using two control CD4⁺8⁺ $\alpha\beta$ TCR⁺ T-cell clones, BK6 and 3C8, which were derived by mitogen stimulation without exposure to *M. tuberculosis* antigens. Previous studies show that BK6 and 3C8 lyse target cell lines expressing CD1a and CD1c, respectively (ref. 1; and S.P., unpublished data). As for all other CD1-reactive T-cell clones described^{1,2,9}, these clones appear to be autoreactive and recognize their nonpolymorphic CD1 ligands in the absence of exogenous antigens. As expected, clones BK6 and 3C8 lysed only C1R transfectants expressing CD1a or CD1c, respectively, and this was not significantly affected by prior incubation of the target cells with *M. tuberculosis* (Fig. 3c, d).

The lack of MHC restriction demonstrated by the preceding experiments argued that MHC-encoded antigen-presenting molecules were not involved in the CD1b-restricted presentation of *M. tuberculosis* antigens to line DN1. As a more stringent test of this hypothesis we produced CD1b transfectants of the T2 cell line, in which extensive chromosomal deletions in both MHC loci result in a complete lack of MHC class II molecule expression^{10,11}. The recently described MHC-linked transporter genes TAP-1 and TAP-2 (reviewed in ref. 12) are also deleted

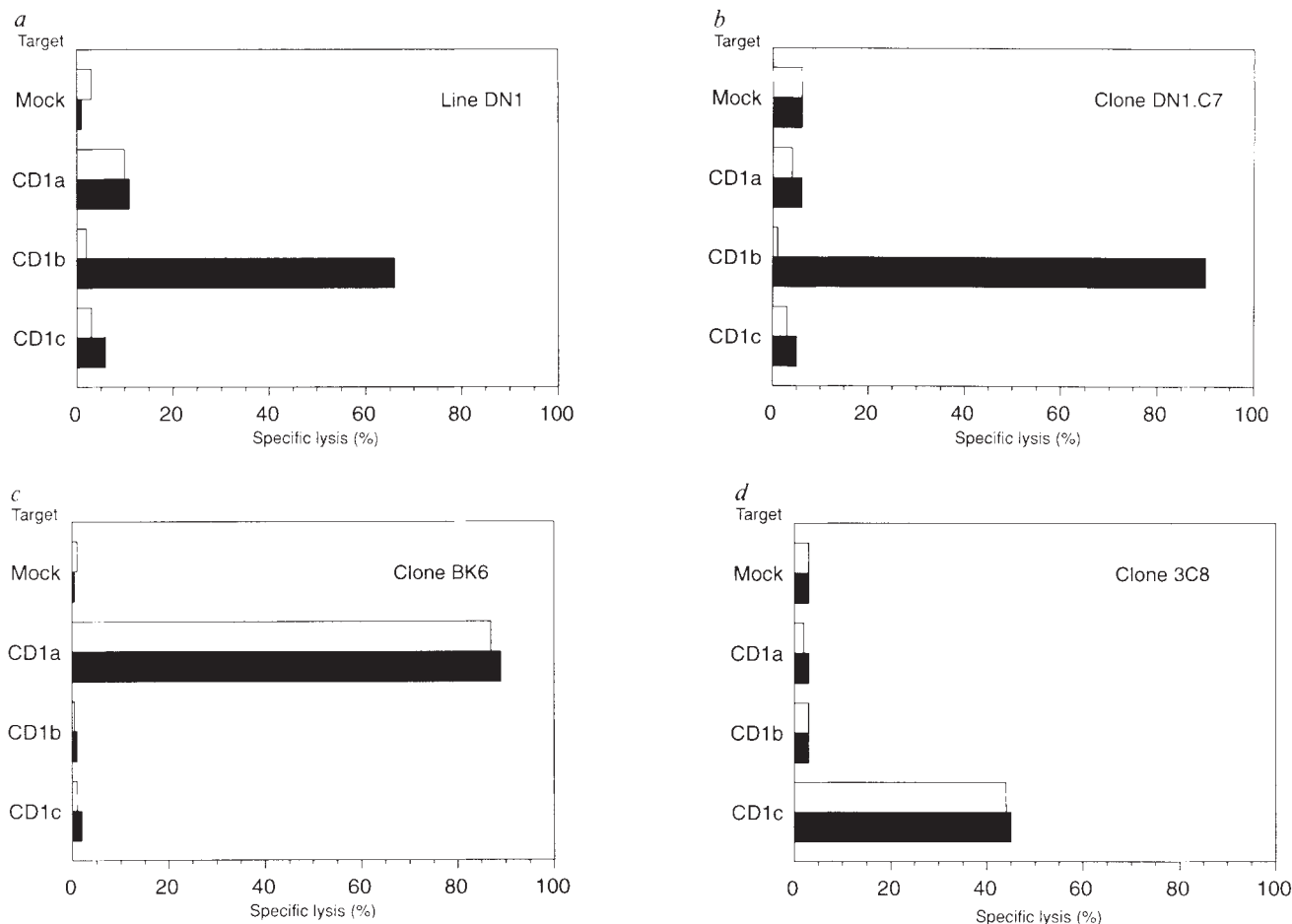


FIG. 3 Presentation of *M. tuberculosis* by CD1 transfectants of B lymphoblastoid cell line C1R. C1R cells stably transfected with vector pSR α -NEO DNA (mock) or with constructs of pSR α -NEO containing complementary DNAs encoding the indicated form of CD1 were cultured for 12 h in medium alone (open bars) or in medium containing *M. tuberculosis* (25 µg protein per ml, solid bars), labelled with ⁵¹Cr and used as target cells for cytolytic assay with *M. tuberculosis* specific T-cell line DN1 (a) or its subclone DN1.C7 (b),

or with autoreactive clones BK6 (c) and 3C8 (d). Effector to target cell ratio was 50:1.

METHODS. Transfection of C1R cells⁹ and assay of specific cytolytic activity by ⁵¹Cr release²⁹ have been described. Clone BK6 was isolated from the blood of a patient with systemic lupus erythematosus as previously described¹, and clone 3C8 was isolated from the blood of a normal donor using the same method.

from T2, resulting in defective expression and function of MHC class II molecules^{13,14}. Nevertheless, transfection of CD1b into T2 led to expression of CD1b on the cell surface at a level similar to that seen on other transfected B-cell lines (S.P., unpublished data) and generated a target cell that presented *M. tuberculosis* to line DN1 (Fig. 4a).

The presentation of exogenous antigens to T cells generally requires uptake and processing of complex protein antigen molecules by antigen-presenting cells, a process which is blocked by aldehyde fixation of the APC surface and by lysosomotropic amines such as chloroquine^{15,16}. By these criteria, CD1b-restricted presentation of *M. tuberculosis* also showed a requirement for antigen uptake and processing. Mild fixation of CD1b⁺ APCs with glutaraldehyde completely abrogated their ability to stimulate line DN1 in the presence of *M. tuberculosis* soluble antigens, although the same APCs pulsed with *M. tuberculosis* before fixation retained their ability to stimulate a proliferative response (Fig. 4b). Furthermore, the presentation of *M. tuberculosis* antigens to line DN1 was strongly inhibited by chloroquine with a dose dependence virtually identical to that for the inhibition of MHC class II-mediated presentation of mycobacterial antigens (Fig. 4c), indicating that processing of antigens for CD1b and MHC class II-restricted responses may

involve similar pathways. T2 cells have recently been shown to be defective in the processing of antigens presented by MHC class II molecules¹⁷. Thus, our finding that CD1b-transfected T2 cells can present *M. tuberculosis* to DN1 suggests that the antigen processing requirements for CD1b and MHC class II molecules, although similar, may not be identical.

We and others have proposed that T cells lacking expression of both CD4 and CD8 molecules may frequently recognize antigens presented by cell surface molecules other than those encoded by classical MHC class I and II loci^{6,18,19}. Here we have shown that one member of the CD1 family, CD1b, could restrict the specific response of MHC-unrestricted CD4⁺CD8⁻ T cells to an exogenous foreign antigen. Like other CD1 proteins, CD1b heavy chains associate noncovalently with β_2 -microglobulin²⁰ and show limited but significant sequence homology to both MHC class I and class II molecules^{3,4}. These structural features of CD1b together with its critical role shown here in antigen recognition support the conclusion that CD1b, and possibly other members of the CD1 family, are nonpolymorphic antigen-presenting molecules encoded by a genetic locus unlinked to the MHC.

Our studies have not yet addressed the possible *in vivo* role of antigen presentation by CD1b or other members of the CD1

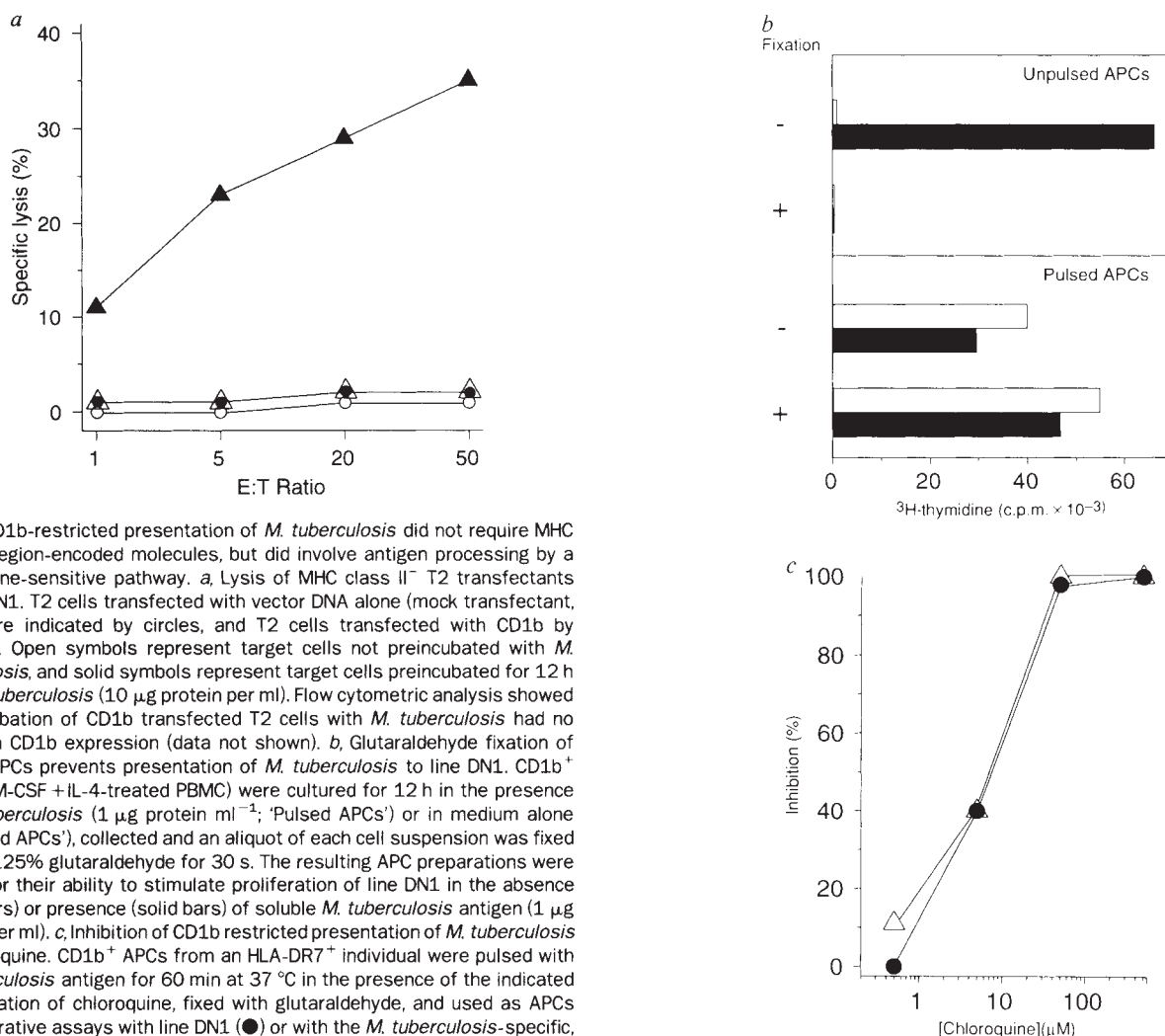


FIG. 4 CD1b-restricted presentation of *M. tuberculosis* did not require MHC class II region-encoded molecules, but did involve antigen processing by a chloroquine-sensitive pathway. **a**, Lysis of MHC class II⁻ T2 transfectants by line DN1. T2 cells transfected with vector DNA alone (mock transfectant, CD1⁻) are indicated by circles, and T2 cells transfected with CD1b by triangles. Open symbols represent target cells not preincubated with *M. tuberculosis*, and solid symbols represent target cells preincubated for 12 h with *M. tuberculosis* (10 μg protein per ml). Flow cytometric analysis showed that incubation of CD1b transfected T2 cells with *M. tuberculosis* had no effect on CD1b expression (data not shown). **b**, Glutaraldehyde fixation of CD1b⁺ APCs prevents presentation of *M. tuberculosis* to line DN1. CD1b⁺ APCs (GM-CSF + IL-4-treated PBMC) were cultured for 12 h in the presence of *M. tuberculosis* (1 μg protein ml⁻¹; 'Pulsed APCs') or in medium alone ('Unpulsed APCs'), collected and an aliquot of each cell suspension was fixed with 0.0125% glutaraldehyde for 30 s. The resulting APC preparations were tested for their ability to stimulate proliferation of line DN1 in the absence (open bars) or presence (solid bars) of soluble *M. tuberculosis* antigen (1 μg protein per ml). **c**, Inhibition of CD1b restricted presentation of *M. tuberculosis* by chloroquine. CD1b⁺ APCs from an HLA-DR7⁺ individual were pulsed with *M. tuberculosis* antigen for 60 min at 37 °C in the presence of the indicated concentration of chloroquine, fixed with glutaraldehyde, and used as APCs in proliferative assays with line DN1 (●) or with the *M. tuberculosis*-specific, HLA-DR7-restricted CD4⁺ T cell line DG.1 (Δ). Results are expressed as per cent inhibition of responses compared to fixed APCs pulsed with *M. tuberculosis* in the absence of chloroquine, and are representative of three similar experiments.

METHODS. Stable transfectants of T2 cells were prepared using the method described for C1R cells⁹. APCs for glutaraldehyde fixation and chloroquine experiments were GM-CSF + IL-4-treated peripheral blood mononuclear cells as described in legend to Fig. 2, and glutaraldehyde fixation and chloroquine

treatment of APCs were done according to published methods^{16,32}. CD4⁺ T-cell line DG.1 was derived from an HLA-DR7⁺ rheumatoid arthritis patient by repeated stimulation of synovial fluid lymphocytes with autologous EBV transformed B cells and *M. tuberculosis* purified protein derivative (PPD, Statens Serum Institute; C.T.M., unpublished data). Proliferation assays were done as in legend to Fig. 2, except that 2×10^5 APCs were added per well and ³H-thymidine incorporation was determined after three days.

family. Although the precise frequency of CD1-restricted T cells is unknown, studies now in progress indicate that *M. tuberculosis*-specific T cells similar to line DN1 can be isolated in reproducible fashion from the blood of other normal donors (S.P., unpublished data). Thus, a potential role for CD1-restricted T cells in normal host defense against microbial disease is clearly inferred. These results suggest a functional parallel between CD1 and MHC class II molecules, because both mediate presentation of exogenous antigens processed through a chloroquine-sensitive pathway, and both can also act as ligands for the TCRs of autoreactive T cells^{1,21}. The limited tissue distribution of CD1 molecules *in vivo* provides a further similarity with the MHC class II family because members of both families are prominently expressed on cell types involved in antigen presentation to T cells, including Langerhan's cells, dendritic cells in lymphoid and many other tissues, B cells and possibly cytokine-activated monocytes (reviewed in ref. 6). In contrast, the lack of structural polymorphism of CD1 molecules, their unique cytokine regulation on monocytes, and the CD4⁺8⁺ phenotype of most CD1-restricted T cells characterized to date are important differences that distinguish the CD1 and MHC antigen-presenting systems. These differences point to a distinct but as yet undefined role for CD1-restricted T cells in cell mediated immunity. □

5. Strominger, J. L. *Cell* **57**, 895-898 (1989).
6. Porcelli, S., Band, H. & Brenner, M. B. *Immun. Rev.* **120**, 137-183 (1991).
7. Bousmell, L. in *Leukocyte Typing IV* (eds Knapp, W. et al.) 251-254 (1989).
8. Zemmour, J., Little, A.-M., Schendel, D. J. & Parham, P. *J. Immun.* **148**, 1941-1948 (1992).
9. Balk, S. P. et al. *Science* **253**, 1411-1415 (1991).
10. Salter, R. D., Howell, D. N. & Cresswell, P. *Immunogenetics* **21**, 235-246 (1985).
11. Erlich, H., Lee, J. S., Petersen, J. W., Bugawan, T. & DeMars, R. *Hum. Immun.* **16**, 205-219 (1986).
12. Parham, P. *Nature* **357**, 193-194 (1992).
13. Hosken, N. A. & Bevan, M. J. *Science* **248**, 367-370 (1990).
14. Wei, M. & Cresswell, P. *Nature* **356**, 443-446 (1992).
15. Ziegler, H. K. & Uhanue, E. R. *Proc. natn. Acad. Sci. U.S.A.* **79**, 175-179 (1982).
16. Chesnut, R. W., Colon, S. M. & Grey, H. M. *J. Immun.* **129**, 2382-2388 (1982).
17. Riberdy, J. M. & Cresswell, P. *J. Immun.* **148**, 2586-2590 (1992).
18. Janeway, C. A. Jr, Jones, B. & Hayday, A. C. *Immun. Today* **6**, 73-76 (1988).
19. Bluestone, J. A. & Matis, L. A. *J. Immun.* **142**, 1785-1788 (1989).
20. Olive, D., Dubreuil, P. & Mawas, C. *Immunogenetics* **20**, 253-264 (1984).
21. Glimcher, L. H. & Shevach, E. M. *J. exp. Med.* **156**, 640-645 (1982).
22. Panchamoorthy, G. et al. *J. Immun.* **147**, 3360-3369 (1991).
23. Reinherz, E. L., Kung, P. C., Goldstein, G., Levey, R. H. & Schlossman, S. F. *Proc. natn. Acad. Sci. U.S.A.* **77**, 1588-1592 (1980).
24. Martin, L. H., Calabi, F., Lefebvre, F.-A., Bilsland, C. A. G. & Milstein, C. *Proc. natn. Acad. Sci. U.S.A.* **84**, 9189-9193 (1987).
25. Brodsky, F. M. & Parham, P. P. *J. Immun.* **128**, 129-135 (1982).
26. Lanier, L. L., Ruitenberg, J. J., Allison, J. P. & Weiss, A. in *Leukocyte Typing III* (ed. McMichael, A. J.) 175-178 (Oxford Univ. Press, Oxford, 1987).
27. Shiue, L., Gorman, S. D. & Parnes, J. R. *J. exp. Med.* **168**, 1993-2005 (1988).
28. Anegón, I., Grolleau, D. & Soullou, J. P. *J. Immun.* **147**, 3973-3980 (1991).
29. Morita, C. T. et al. *Eur. J. Immun.* **21**, 2999-3007 (1991).
30. Favoloro, E. J. et al. *Disease Markers* **4**, 261-270 (1986).
31. Shaw, S., Ziegler, A. & DeMars, R. *Hum. Immun.* **12**, 191-211 (1985).
32. Roncarolo, M. G. et al. *J. Immun.* **147**, 781-787 (1991).
33. Albertson, D. G., Fishpool, R., Sherrington, P., Nacheva, E. & Milstein, C. *EMBO J.* **7**, 2801-2805 (1988).
34. Martin, L. H., Calabi, F. & Milstein, C. *Proc. natn. Acad. Sci. U.S.A.* **83**, 9154-9158 (1986).
35. Calabi, F., Jarvis, J. M., Martin, L. & Milstein, C. *Eur. J. Immun.* **19**, 285-292 (1989).

Received 16 June; accepted 16 October 1992.

1. Porcelli, S. et al. *Nature* **341**, 447-450 (1989).
2. Faure, F., Jitsukawa, S., Miossec, C. & Hercend, T. *Eur. J. Immun.* **20**, 703-706 (1990).
3. Calabi, F. & Milstein, C. *Nature* **323**, 540-543 (1986).
4. Balk, S. P., Bleicher, P. A. & Terhorst, C. *Proc. natn. Acad. Sci. U.S.A.* **86**, 252-256 (1989).

ACKNOWLEDGEMENTS. Recombinant human cytokines were generously provided by Genetics Institute (Cambridge, MA; GM-CSF), Schering Corp. (Bloomfield, NJ; IL-4 and IFN γ) and the Ajinomoto Company (Kawasaki, Japan; IL-2). We thank P. Cresswell for the C1R and T2 cell lines, P. Brennan for *M. leprae* bacilli and the following for providing antibodies: E. Favoloro (WM-25), W. Knapp (10C3), D. Olive and C. Mawas (4A7.6), R. Kurrel (BMA-031) and E. Reinherz (25T8-5H7). This work was supported by the NIH, the Arthritis Foundation and the Leukemia Society of America.

Ornithine decarboxylase is degraded by the 26S proteasome without ubiquitination

Yasuko Murakami, Senya Matsufuji, Takaaki Kameji, Shin-ichi Hayashi, Kazuei Igarashi*, Tomohiro Tamura†, Keiji Tanaka† & Akira Ichihara†

Department of Nutrition, The Jikei University School of Medicine, Minato-ku, Tokyo 105, Japan

* Faculty of Pharmaceutical Sciences, Chiba University, Yayoicho, Chiba 260, Japan

† Institute for Enzyme Research, The University of Tokushima, Tokushima 770, Japan

ORNITHINE decarboxylase (ODC), a key enzyme in polyamine biosynthesis, is the most rapidly turned over mammalian enzyme¹. We have shown that its degradation is accelerated by ODC antizyme^{2,3}, an inhibitory protein induced by polyamines^{4,5}. This is a new type of enzyme regulation and may be a model for selective protein degradation. Here we report the identification of the protease responsible for ODC degradation. Using a cell-free degradation system, we demonstrate that immunodepletion of proteasomes from cell extracts causes almost complete loss of ATP- and antizyme-dependent degradation of ODC. In addition, purified 26S proteasome complex, but not the 20S proteasome, catalyses ODC degradation in the absence of ubiquitin. These results strongly suggest that the 26S proteasome, widely viewed as specific for ubiquitin-conjugated proteins, is the main enzyme responsible for ODC degradation. The 26S proteasome may therefore have a second role in ubiquitin-independent proteolysis.

We first investigated whether purified exogenous ODC is degraded by cell extracts. As shown in Fig. 1a, added ODC is rapidly degraded in the presence of both ATP and antizyme, in essentially the same way as endogenous ODC in extracts from CHO cells² or HTC cells (data not shown). Degradation of

ODC in both HTC and CHO cell extracts is strongly inhibited by haemin (Fig. 1b). Haemin blocks ATP-dependent degradation of ubiquitin (Ub)-protein conjugates but has less effect on the ligation of Ub to proteins, resulting in accumulation of Ub-protein conjugates⁶. Fluorograms of preparations with or without haemin showed no detectable bands of ³⁵S-labelled ODC of high molecular weight, which would correspond to ubiquitinated forms of the enzyme, even after long exposures, suggesting that ODC degradation is not mediated by Ub. This is consistent with previous findings that ODC degradation in reticulocyte lysates is not affected by immunodepletion of E1, the Ub-activating enzyme⁷, nor by incubation at the non-permissive temperature of ts85, a mouse cell line with a thermolabile E1 (ref. 8). Neither of the latter studies was conclusive, however, because ODC degradation was slow in the reticulocyte lysate in the absence of antizyme and there is a report of persistent Ub conjugation in non-permissive ts85 cells⁹.

To define the pathway of ODC degradation further, we tested the effects of nucleotides and various other compounds on the degradation (Table 1). Like ATP, all nucleotide triphosphates stimulated degradation of ODC, indicating a broad nucleotide specificity; ADP was also stimulatory, probably by virtue of its conversion to ATP by adenylate kinase, whereas AMP and non-hydrolysable ATP analogues had no effect. The magnesium requirement indicates that ATP hydrolysis is necessary for the degradation of ODC, as does the block of ODC degradation induced by vanadate, an ATPase inhibitor. ODC degradation is also sensitive to *N*-ethylmaleimide, a sulphhydryl reagent. High levels of EGTA (5 mM) and E64 (0.1 mM), which both inhibit calpain, did not prevent proteolysis, neither did peptide inhibitors of lysosomal proteases such as leupeptin, antipain, chymostatin and pepstatin at 0.1 mM or inhibitors of serine proteases, like phenylmethanesulphonylfluoride (0.5 mM), 3,4-dichloroisocoumarin (0.2 mM) and soybean trypsin inhibitor (10 μ g ml⁻¹), indicating that the protease responsible for ODC degradation is neither a lysosomal enzyme nor calpain.

The proteasome, which is a multicatalytic protease, catalyses an energy-dependent, non-lysosomal proteolytic pathway¹⁰⁻¹².

To determine whether the proteasome is involved in ODC degradation, we examined the effect of immunoprecipitation of proteasome from HTC cell extracts on the degradation activity against both ODC and a fluorogenic peptide Suc-LLVY-MCA (succinyl-Leu-Leu-Val-Tyr-4-methylcoumaryl-7-amide), which is a good substrate for the proteasome¹³. Addition of anti-proteasome antibodies gave a similar dose-dependent decrease in the degradation of both substrates, with high concentrations causing almost complete loss of the ATP- and antizyme-dependent ODC degrading activity; control antibodies had no effect (Fig. 2). These results strongly suggest that the proteasome is necessary for ODC degradation.

The proteasome was initially identified as a 20S complex compound of many subunits (M_r 21–31k), but recent studies have revealed that cells actually contain a 26S proteasome complex^{14–17}, consisting of a 20S proteasome with many associated proteins (M_r 35–110K; see Fig. 3a, inset). The 26S complex catalyses energy-dependent degradation of ubiquitinated proteins. To see which type of proteasome was responsible for ODC degradation, we first prepared the 26S proteasome complex. Figure 3a shows the profile on a glycerol density gradient of a rat liver 26S proteasome preparation at the final purification step¹⁸. The profile of ODC degradation activity coincided with the activity of the 26S proteasome complex as determined by both fluorogenic peptide hydrolysis and ATP-dependent degradation of ¹²⁵I-labelled lysozyme-Ub conjugates. The intrinsic ATPase activity of the 26S proteasome complex¹⁸ gave the same profile (data not shown).

We next examined ODC degradation by purified 26S and 20S proteasome complexes. As shown in Fig. 3b, degradation of ODC by the purified 26S proteasome, which did not contain Ub or Ub-conjugating enzymes, was similar to degradation by crude cell extracts: it required ATP hydrolysis and was antizyme-dependent. Inhibition of ATPases by vanadate led to marked suppression of ODC-degrading activity, suggesting that the intrinsic ATPase of the 26S complex^{18,19} supplied the energy for ODC degradation. As with crude cell extract (Fig. 1b), no intermediate bands were detected by SDS-PAGE analysis of the assay mixture containing the purified 26S complex (Fig. 3b, inset), indicating rapid and complete proteolysis. In contrast, the purified 20S proteasome did not degrade ODC in the presence of ATP and antizyme, even when added in large amounts (Fig. 3b).

These findings show that ODC is degraded specifically by the 26S proteasome in an ATP- and antizyme-dependent manner. To our knowledge, this is the first demonstration that a specific protease degrades a specific enzyme *in vitro* in an ATP-dependent manner.

TABLE 1 Properties of energy- and antizyme-dependent degradation of ornithine decarboxylase in cell extracts

Nucleotide	ODC degradation	
	CHO cells	HTC cells
None	16	8
ATP	52	61
CTP	55	64
UTP	53	52
GTP	52	55
ADP	53	62
AMP	12	7
α , β -methylene ATP	13	4
β , γ -methylene ATP	20	11
ATP, no Mg^{2+}	27	32
ATP, no Mg^{2+} , +EDTA (5 mM)	8	6

Compound	Concentration (μ M)	ODC degradation	
		CHO cells	HTC cells
None		100	100
Haemin	250	33	21
N-ethylmaleimide	5	0	ND
Vanadate	50	19	15
	100	23	18

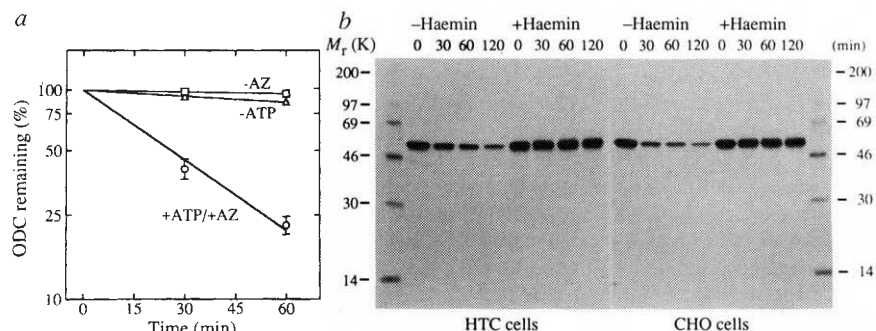
In the upper part of the table, the effects of different nucleotides on ODC degradation are shown. ³⁵S-labelled ODC was incubated for 30 min with cell extracts in the presence of antizyme as for Fig. 1, except that the ATP regeneration system was omitted and the indicated nucleotide (5 mM) was added instead of ATP. Mg^{2+} was included unless otherwise indicated. In the lower part of the table, degradation of ³⁵S-labelled ODC, assayed as for Fig. 1, in the presence of different compounds is shown. ATP, $MgCl_2$ and antizyme were included in the assay. Values give degradation in the presence of test compounds as a percentage of the control value. ND, not determined.

Finally, we tested whether antizyme affects the ATP-dependent degradation of other proteins or Ub conjugates by the 26S proteasome complex. Our preparation of the 26S proteasome complex did not degrade free proteins, such as ¹²⁵I-labelled casein and ¹²⁵I-labelled lysozyme, in contrast to previous reports^{10,12,16}. Moreover, antizyme did not elicit ATP-dependent degradation of these proteins, nor did it stimulate degradation of Ub-conjugated lysozyme by the 26S proteasome (data not shown). These results exclude the possibility that antizyme is a general regulator of the 26S proteasome and are consistent with a report that specific binding of antizyme to ODC is required for acceleration of ODC degradation by polyamines²⁰. There is no sequence similarity between antizyme and Ub.

How does the 26S proteasome recognize the ODC-antizyme complex? It has been suggested that ODC contains structural

FIG. 1 Degradation of purified ODC by cell extracts. *a*, ATP- and antizyme (AZ)-dependent degradation of ODC. ³⁵S-ODC was incubated with a CHO cell extract in the presence of ATP (□), AZ (△) or both (○). The amount of remaining ODC was measured at the indicated times and is expressed as a percentage of the starting amount. Points and bars indicate means $\pm$ s.d. for 3 experiments. *b*, Inhibition by haemin (250 μ M) of ODC degradation by extracts of HTC and CHO cells. Note that no higher molecular weight bands of Ub-ODC conjugates were detected in either the presence or absence of haemin.

METHODS. Substrate ODC was prepared from ODC-overproducing EXOD-1 cells derived from mouse FM3A cells³⁰. Cells were cultured with α -difluoromethylornithine, an irreversible inhibitor of ODC, which did not affect degradation of ODC²⁸ but induced the production of a large amount of the enzyme. ODC was metabolically labelled in cells with ³⁵S-methionine and ³⁵S-cysteine for 2.5 h, then purified by immunoaffinity chromatography²⁹. Extracts of CHO cells (DF3) and HTC cells and purified recombinant antizyme were prepared as described². In *a*, ³⁵S-ODC (5.2 ng) was incubated with the cell extract



supplemented with antizyme (4.9 ng) and an ATP-regenerating system at 37 °C for 30–60 min in a volume of 25 μ l made up with degradation assay mixture². At the indicated times, the amount of trichloroacetic acid-insoluble radiolabel was determined by precipitation. In *b*, samples were quenched at the indicated times with sample buffer, and analysed by SDS-PAGE and fluorography.

FIG. 2 Effect of immunoprecipitation of the proteasome on ATP- and antizyme-dependent degradation of ODC by an HTC cell extract. An HTC cell extract was treated with anti-proteasome IgG (●) or control IgG (○), and the degradation activity of the supernatants assayed using Suc-LLVY-MCA (a) or 35 S-ODC (b) as substrate.

METHODS. An HTC cell extract (200 μ g protein) was incubated in a total volume of 210 μ l at 4°C for 2h with the indicated amounts of anti-proteasome IgG¹³ or control IgG, and then centrifuged at 14,000g for 30 min. Supernatants (5 μ l) were used for proteolytic assay; ODC degradation was assayed in the presence of ATP and antizyme as for Fig. 1a. Suc-LLVY-MCA degradation was measured as reported¹⁸. Values are shown as percentages of the activity measured with no added IgG, namely 4.2 nmol h⁻¹ for peptide hydrolysis and 60% per 30 min for ODC degradation.

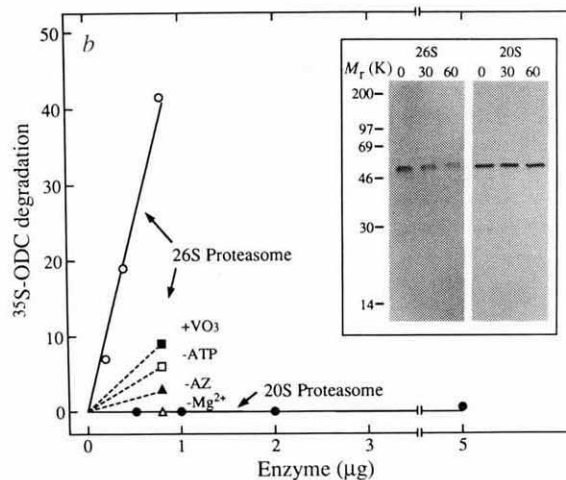
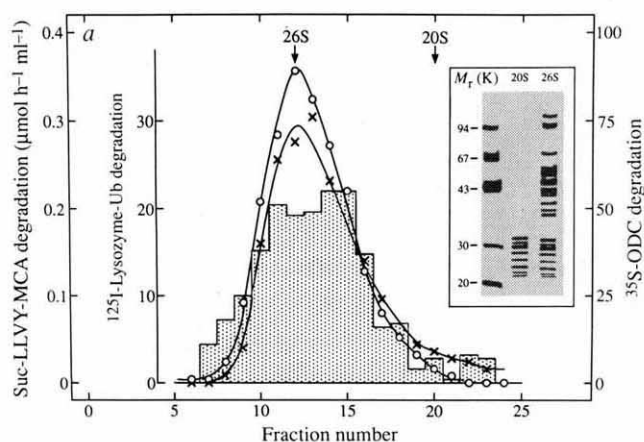
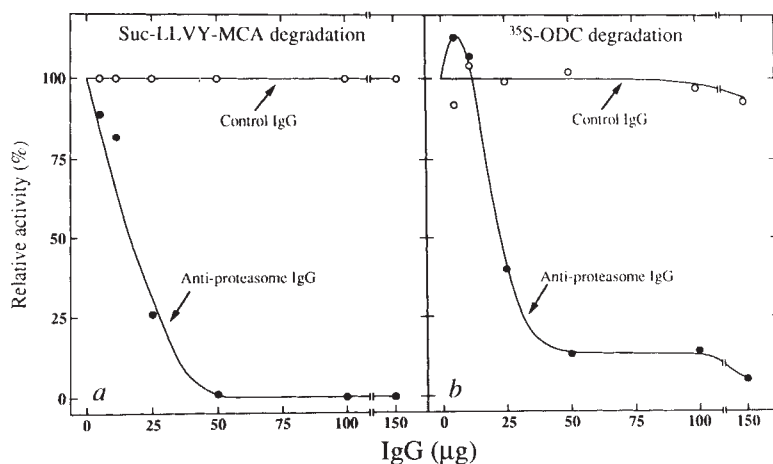


FIG. 3 ODC degradation by the 26S proteasome complex. a, ODC degradation (per cent per 40 min) by the 26S proteasome from rat liver separated by glycerol density-gradient centrifugation (1 ml fractions). Degradations of 35 S-ODC (shaded histograms), 125 I-labelled lysyl-ubiquitin conjugates (○; per cent per 60 min) and Suc-LLVY-MCA (×) are shown. Arrows indicate elution positions of the 20S and 26S proteasomes. The inset shows SDS-PAGE of 20S and 26S proteasomes purified from rat liver. b, Per cent ODC degradation (in the presence of ATP and antizyme, in 30-min assay) by the purified 26S proteasome (○) but not the 20S proteasome (●) from rat liver. ODC

degradation by the 26S proteasome was also assayed with 100 μ M vanadate (VO_3^- ; ■) or without ATP (□), antizyme (AZ, ▲), or MgCl_2 (△). The inset shows fluorograms of 35 S-ODC degradation by 20S and 26S proteasomes (0, 30 and 60 min).

METHODS. ODC degradation activity was assayed without the ATP regeneration system. In a, samples of 5 μ l, 10 μ l and 20 μ l, respectively, of each fraction were assayed for ATP- and antizyme-dependent ODC degradation (for 40 min), and degradation of peptide (10 min) and lysyl-ubiquitin conjugate (2 h) as described¹⁸.

domains responsible for its rapid degradation^{21–24}, including 'PEST' sequences which are common in short-lived proteins²⁵. These regions of ODC might be exposed on binding to antizyme and be recognized by the 26S proteasome. An interesting possibility is that antizyme-stimulated ODC degradation may be similar to other degradation control mechanisms, such as the stimulation of degradation of p53 by binding to the E6 protein of oncogenic human papilloma virus²⁶ and of the degradation of abnormal proteins by binding to the chaperonin dnaK²⁷.

It was believed that the 26S proteasome complex is specific for degradation of ubiquitinated proteins^{14–18}, but our finding that it also rapidly degrades the non-ubiquitinated ODC-antizyme complex suggests that it may have a more general function in selective removal of short-lived proteins by recognizing degradation signals other than ubiquitination. Alternatively, ODC may be degraded by a subspecies of the 26S complex that does not recognize ubiquitinated proteins but is co-purified with the ubiquitination-specific 26S proteasome. □

Received 1 June; accepted 21 October 1992.

- Pegg, A. E. *Biochem. J.* **234**, 249–262 (1986).
- Murakami, Y., Tanaka, K., Matsufuji, S., Miyazaki, Y. & Hayashi, S. *Biochem. J.* **283**, 661–664 (1992).
- Murakami, Y., Matsufuji, S., Miyazaki, Y. & Hayashi, S. *J. Biol. Chem.* **267**, 13138–13141 (1992).
- Fong, W. F., Heller, J. S. & Canellakis, E. S. *Biochim. biophys. Acta* **428**, 456–465 (1976).
- Hayashi, S. & Canellakis, E. S. in *Ornithine Decarboxylase: Biology, Enzymology, and Molecular Genetics* (ed. Hayashi, S.) 47–58 (Pergamon, New York and Oxford, 1989).
- Haas, A. L. & Rose, I. A. *Proc. natn. Acad. Sci. U.S.A.* **78**, 6845–6848 (1981).
- Bercovich, Z., Rosenberg-Hasson, Y., Ciechanover, A. & Kahana, C. *J. Biol. Chem.* **264**, 15949–15952 (1989).
- Rosenberg-Hasson, Y., Bercovich, Z., Ciechanover, A. & Kahana, C. *Eur. J. Biochem.* **185**, 469–474 (1989).
- Loetscher, P., Pratt, G. & Rechsteiner, M. *J. Biol. Chem.* **266**, 11213–11220 (1991).
- Goldberg, A. L. *Eur. J. Biochem.* **205**, 9–23 (1992).
- Tanaka, K., Tamura, T., Yoshimura, T. & Ichihara, A. *The New Biologist* **4**, 173–187 (1992).
- Goldberg, A. L. & Rock, K. L. *Nature* **357**, 375–379 (1992).
- Tanaka, K. et al. *J. Biol. Chem.* **263**, 16209–16217 (1988).
- Hough, R., Pratt, G. & Rechsteiner, M. *J. Biol. Chem.* **262**, 8303–8313 (1987).

- Eytan, E., Ganoh, D., Armon, T. & Herskko, A. *Proc. natn. Acad. Sci. U.S.A.* **86**, 7751–7755 (1989).
- Driscoll, J. & Goldberg, A. L. *J. Biol. Chem.* **265**, 4789–4792 (1990).
- Orino, E. et al. *FEBS Lett.* **284**, 206–210 (1991).
- Kanayama, H. et al. *Eur. J. Biochem.* **206**, 567–578 (1992).
- Armon, T., Ganoh, D. & Herskko, A. *J. Biol. Chem.* **265**, 20723–20725 (1990).
- Li, X. & Coffino, P. *Molec. cell. Biol.* **12**, 3356–3362 (1992).
- Ghoda, L., van Daalen Wetters, T., Macrae, M., Ascherman, D. & Coffino, P. *Science* **243**, 1493–1495 (1989).
- Rosenberg-Hasson, Y., Bercovich, Z. & Kahana, C. *Eur. J. Biochem.* **196**, 647–651 (1991).
- Lu, L., Stanley, B. A. & Pegg, A. E. *Biochem. J.* **277**, 671–675 (1991).
- Ghoda, L., Sidney, D., Macrae, M. & Coffino, P. *Molec. cell. Biol.* **12**, 2178–2185 (1992).
- Rogers, S., Wells, R. & Rechsteiner, M. *Science* **234**, 364–368 (1986).
- Scheffner, M., Werner, B. A., Hultberg, J. M., Levine, A. J. & Howley, P. M. *Cell* **63**, 1129–1136 (1990).
- Sherman, M. & Goldberg, A. L. *EMBO J.* **11**, 71–77 (1992).
- Seely, J. E., Poso, H. & Pegg, A. E. *J. Biol. Chem.* **257**, 7549–7553 (1982).
- Nishiyama, M. et al. *Prep. Biochem.* **18**, 227–238 (1988).
- Kameji, T., Hayashi, S., Hoshino, K., Kakinuma, Y. & Igarashi, K. *Biochem. J.* (in the press).

Raf functions downstream of Ras1 in the Sevenless signal transduction pathway

Barry Dickson, Frank Sprenger^{*†}, Deborah Morrison[‡] & Ernst Hafen

Zoologisches Institut der Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

^{*} Max Planck Institut für Entwicklungsbiologie, Abteilung Genetik, Spemannstrasse 35, D-7400 Tübingen, Germany

[†] Present address: Department of Biochemistry and Biophysics, University of California, San Francisco, California 94143-0448, USA

[‡] ABL-Basic Research Program, NCI-Frederick Cancer Research and Development Center, PO Box B, Frederick, Maryland 21702-1201, USA

SPECIFICATION of the R7 cell fate in the developing *Drosophila* eye requires activation of the Sevenless (Sev) receptor tyrosine kinase, located on the surface of the R7 precursor cell, by its interaction with the Boss protein, expressed on the surface of the neighbouring R8 cell¹⁻³. Four genes that participate in the intracellular transmission of this signal have so far been identified and molecularly characterized: *Ras1*, *Sos*, *Gap1* and *sina* (refs 4-8). The *Drosophila* homologue of the mammalian Raf-1 serine/threonine kinase, which has been implicated in signal transduction pathways activated by many receptor tyrosine kinases (reviewed in refs 9 and 10), is encoded by the *raf* locus (also known as *l(1)polehole*¹¹, *Draf-1*¹² or *Draf*¹³). Here we show that the *Drosophila* Raf serine/threonine kinase also plays a crucial role

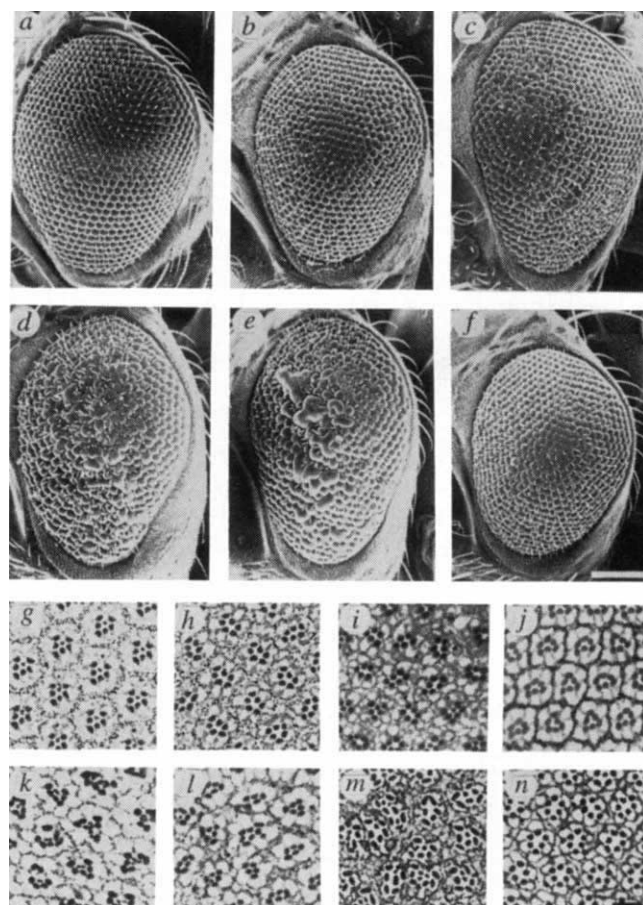
in the R7 pathway: the response to Sev activity is dependent on *raf* function, and a constitutively activated Raf protein can induce R7 cell development in the absence of *sev* function. We also present genetic evidence suggesting that Raf acts downstream of Ras1 and upstream of Sina in this signal transduction cascade.

Maternally derived Raf protein is required during embryogenesis for the response to activation of the *torso* receptor tyrosine kinase¹³. Strong loss-of-function *raf* mutations block cell proliferation^{11,12}, and as anticipated, our attempts to examine the consequences of complete loss of *raf* function during eye development by generating *raf*⁻ clones of cells were unsuccessful. Males hemizygous for the weak *raf*^{HM7} allele do, however, survive to adulthood when reared at 18 °C (ref. 14). These flies have slightly rough eyes (Fig. 1b). Apical sections through such eyes reveal that the R7 cell is missing in most ommatidia and some of the outer photoreceptor cells (R1-6) are also frequently absent (Fig. 1i). Furthermore, we find that this *raf* allele markedly suppresses the rough-eye phenotype produced by a gain-of-function *sev* allele (data not shown). Thus, normal Raf activity is required for efficient signalling by either a wild-type or constitutively activated Sev kinase, suggesting that Raf acts directly in the Sev signal transduction pathway. Additionally, Raf, like Ras1 (ref. 4), may also play a part in the development of the outer photoreceptor cells.

To determine whether Raf activation is sufficient to induce R7 cell development, we expressed wild-type and constitutively activated Raf proteins in all *sev*-expressing cells of the developing eye disc. Constitutive activation of Raf was achieved by fusing the Raf kinase domain to the signal sequence and extracellular and transmembrane domains of the Torso protein^{15,16}. Recombinant fusion and wild-type Raf proteins produced in

FIG. 1 Eye morphology of flies carrying various loss- and gain-of-function *raf* and *Ras1* alleles: a-f, Scanning electron micrographs of eyes of wild-type (a), *raf*^{HM7} (b), *sev*^{d2}, *sE-raf*^{tor} (c), *sev*^{d2}, *sE-raf*^{tor4021} (d), *sev*^{d2}, *sevRas1*^{V12} (e), and *raf*^{HM7}, *sev*^{d2}, *sevRas1*^{V12} (f) flies. g-l, Tangential sections through the eyes of wild-type (g), *sev*^{d2} (h), *raf*^{HM7} (i), *sev*^{d2}, *sE-raf*^{tor} (j), *sev*^{d2}, *sE-raf*^{tor4021} (k), *sev*^{d2}, *sE-raf*^{tor4021}, *sina*³ (l), *sev*^{d2}, *sevRas1*^{V12} (m), and *raf*^{HM7}, *sev*^{d2}, *sevRas1*^{V12} (n) flies. A small number of R7-like cells are observed in *sev*^{d2}, *sE-raf*^{tor4021}, *sina*³ eyes (l), presumably because the *sina*³ allele retains some activity. The flies shown in b, e, f, i, m and n are males reared at 18 °C; all others are females reared at 25 °C. The *sE-raf* transformant shown in j is homozygous for the insertion chromosome, all other transgenic flies carry a single copy of the indicated construct. Flies homozygous for the *sE-raf*^{tor4021} construct (d) display a severe roughening of the eye, comparable to that observed in heterozygotes for the *sE-raf*^{tor4021} construct (d). This roughening correlates with an increase in the number of additional R7-like cells (data not shown). Thus, we believe that the *torso*⁴⁰²¹ domain of the *raf*^{tor4021} hybrid produces a greater degree of activation of the Raf kinase than does the wild-type *torso* domain. The suppression of the rough-eye phenotype of *Ras1*^{V12} by *raf*^{HM7} (e, f) is attributable to a reduction in the number of *Ras1*^{V12}-induced R7 cells (m, n). Scale bars: f, 100 µm; n, 10 µm.

METHODS. The *sE-raf*, *sE-raf*^{tor}, *sE-raf*^{tor4021} and *sE-raf*^{tor4021KV} transformation constructs are pW8 (ref. 28) derivatives containing *sev* enhancer sequences from 6,347 to 7,564 (ref. 1) and *hsp70* promoter sequences from -250 to +90 (ref. 17). *sE-raf* contains the *raf* cDNA DrafG3 (F.S., M. Trosclair and D.M., manuscript in preparation). Repeated heat-shock induction of this construct rescues an amorphic *raf* mutation, confirming that it encodes functional Raf kinase (data not shown). For the activated *raf* constructs, a 3' cDNA fragment starting at position 2,111 of the genomic *raf* sequence¹² was amplified by polymerase chain reaction (PCR), introducing a *Bst*II site at the 5' end and an *Eco*RI site at the 3' end of the amplified fragment. This fragment was cloned into the *Bst*II site (position 2,714 of the genomic *torso* sequence¹⁶) of the cDNAs pBtor or pB4021²⁹. The resulting cDNAs encode proteins containing the first 455 amino acids of Torso, including the extracellular and transmembrane domains, followed by the 411 C-terminal amino acids of Raf. The *torso*⁴⁰²¹ mutation is an A to G transition at position 2,279, resulting in a cysteine-for-tyrosine substitution at amino-acid position 327 (ref. 29). The *sE-raf*^{tor4021KV} construct is identical to *sE-raf*^{tor4021}, except for an AAA to GTC codon change introduced by *in vitro* mutagenesis to substitute a valine residue for lysine at the position corresponding to residue 382 of the Raf protein¹². *Drosophila*



germ-line transformation, scanning electron microscopy and histological preparations were done as described²⁵.

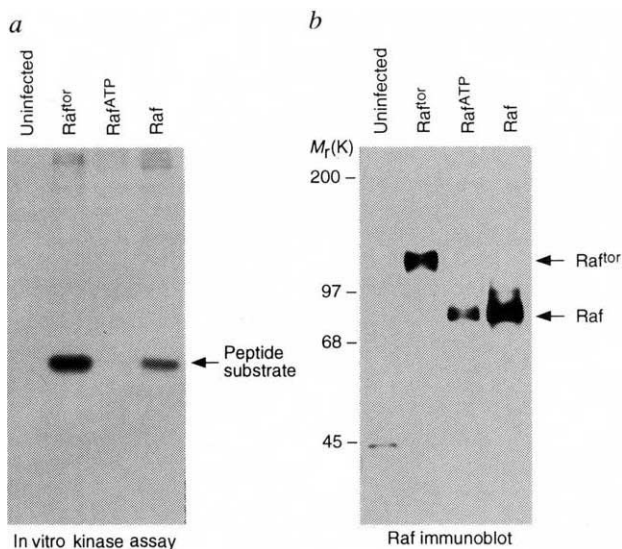


FIG. 2 *a*, Demonstration of the *in vitro* kinase activity of wild-type and fusion Raf proteins. *b*, Recombinant Raf proteins produced in Sf9 cells. METHODS. Lysates were prepared from uninfected Sf9 cells and Sf9 cells expressing *Raf^{tor}*, kinase-inactivated Raf (*Raf^{ATP}*) or wild-type Raf proteins. *a*, Anti-Raf immunoprecipitates were incubated in 50 μ l kinase buffer containing 25 mM HEPES, pH 7.4, 10 mM $MnCl_2$, 1 mM DTT, 10 μ M $[\gamma^{32}P]$ ATP and 2 μ g Raf-1 autophosphorylation peptide as an exogenous substrate. Kinase assays were incubated at 25 $^{\circ}$ C for 20 min. Samples were resolved on 15% Tricine gels and phosphoproteins visualized by autoradiography. *b*, Lysates were resolved on 7.5% SDS-PAGE, transferred to nitrocellulose filters and examined by immunoblot analysis using anti-Raf antibodies. Migration of M_r standards ($\times 10^{-3}$) is shown on the left.

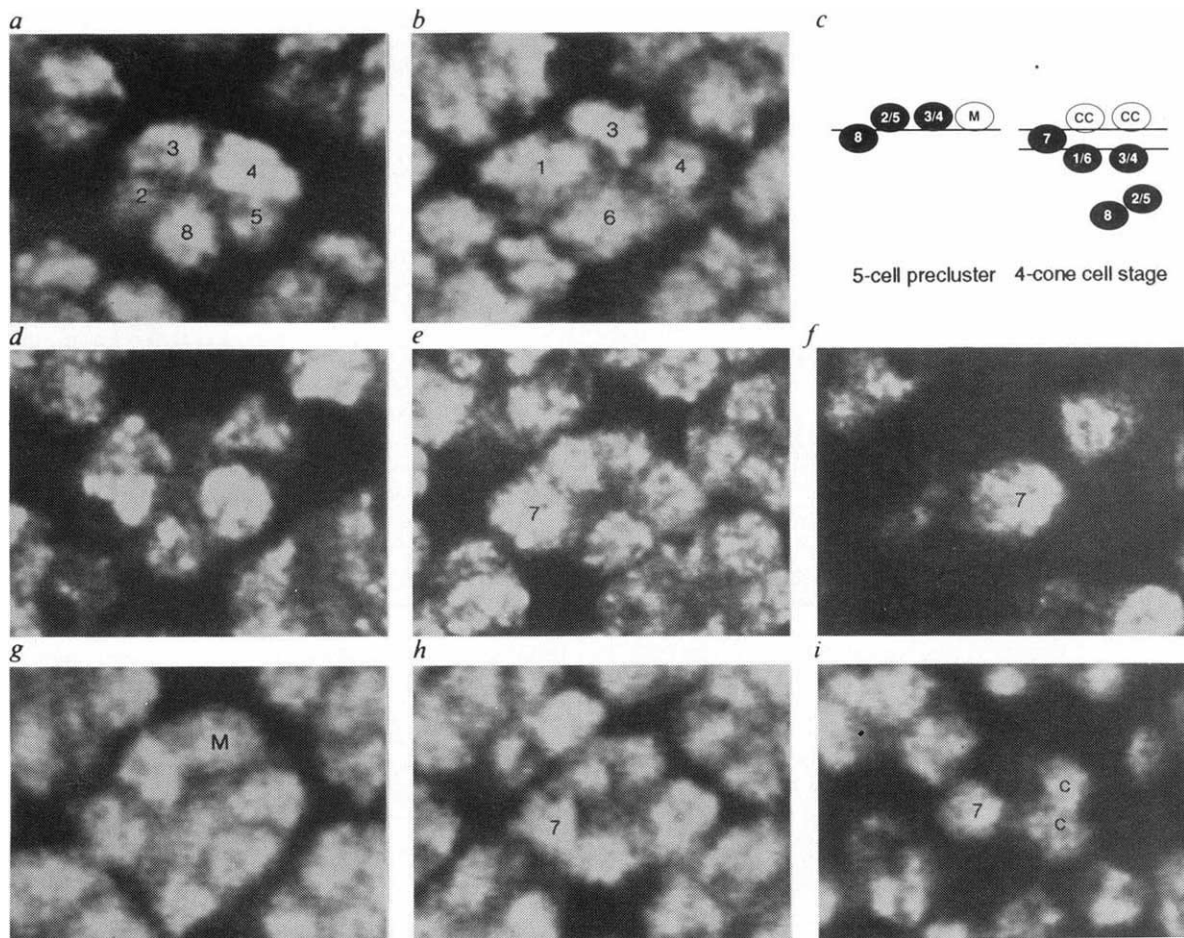


FIG. 3 Ommatidial assembly in *sev^{d2}* (*a, b*), wild-type (*c–e*), and *sev^{d2}; sE-raf^{tor4021}* (*f–h*) flies examined by anti-Elav staining of late 3rd instar larval eye discs. *a, d*, and *g*. At the 5-cell precluster stage, Elav antigen is detected in the nuclei of the R8, R2, R5, R3 and R4 precursors in *sev^{d2}* (*a*) and wild-type (*d*) discs. In *sev^{d2}; sE-raf^{tor4021}* discs, the mystery cell, located between R3 and R4, also expresses Elav (*g*). *b, e, f, h* and *i*. At the four-cone cell stage, Elav staining is observed in the precursors of R1–6 and R8 in each case, and in the R7 precursor in wild-type (*e, f*) and *sev^{d2}; sE-raf^{tor4021}* (*h, i*) discs, but not in *sev^{d2}* discs (*b*). Additionally, the cone cell precursors, which do not express the Elav antigen in wild-type (*f*) or *sev^{d2}* discs, frequently show Elav staining in *sev^{d2}; sE-raf^{tor4021}* discs (*i*).

c, A schematic representation of a lateral view of the 5-cell precluster and four-cone cell stages of ommatidial assembly, showing the relative positions of the various nuclei. Horizontal lines indicate the focal planes shown in the accompanying figures. Numbers 1–8, photoreceptors R1–R8; M, mystery cell; c, cone cell. Scale bar, 5 μ m.

METHODS. Eye discs were stained with a 1:25 dilution of affinity-purified rat polyclonal anti-Elav serum (gift from K. White), followed by an FITC-conjugated secondary antibody (Southern Biotech) as described⁷. Optical sections of whole-mount disc preparations were obtained on an MRC 600 confocal microscope.

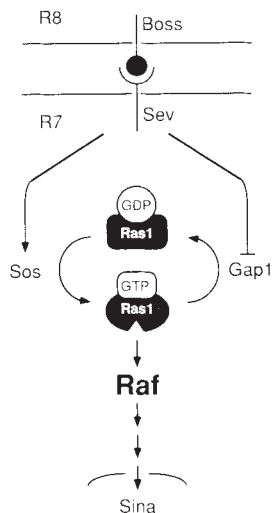


FIG. 4 Model for the role of Raf in the Sevenless signal transduction cascade. Activation of the Sev receptor tyrosine kinase as a result of its interaction with the Boss ligand³ leads to an increase in the amount of active GTP-bound Ras1 protein, either by stimulation of Sos (a guanine-nucleotide-exchange factor homologue)^{4,6} or inhibition of Gap1 (a putative Ras GTPase-activating protein)⁷, or both⁷. Raf acts downstream of Ras1, playing a crucial role in the transduction of this signal towards the nucleus, where one possible target, the Sina protein, is localized⁸.

Sf9 cells phosphorylated a synthetic peptide containing the autophosphorylation sites of the mammalian Raf-1 protein, with the fusion protein showing a raised level of kinase activity compared with the wild-type protein (Fig. 2). These constructs, and an analogous fusion construct containing the extracellular domain of an activated Torso protein²⁹ were then placed under the transcriptional control of the *sev* enhancer (*sE*)¹⁷ and *hsp70* promoter¹⁸ and introduced into the *Drosophila* germ line by P-element-mediated transformation. Expression of wild-type Raf protein under the *sev* enhancer (*sE-raf*) had no effect on eye morphology (data not shown), but a dominant rough-eye phenotype was observed in each of eight independent transgenic lines containing either of the fusion constructs (*sE-raf^{tor}* and *sE-raf^{tor4021}*; Fig. 1c, d). This phenotype requires a functional Raf kinase domain because it was not observed in transformants carrying an analogous construct in which a valine residue replaces the conserved lysine residue in the Raf ATP-binding site (*sE-raf^{tor4021KV}*; data not shown). A mutation at the corresponding site in the *v-raf* oncogene likewise abolishes its transforming potential¹⁹.

Apical sections through the eyes of *sE-raf^{tor}* and *sE-raf^{tor4021}* transformants reveal the presence of multiple R7-like cells, even in a *sev* mutant background (Fig. 1k). In contrast, even two copies of the *sE-raf* construct are insufficient to permit R7 cell development in a *sev* background (Fig. 1j). To determine the origin of the additional photoreceptor cells in *sE-raf^{tor4021}* transformants, we stained eye discs from *sev* and *sev; sE-raf^{tor4021}* larvae with an antibody specific to the neural antigen encoded by the *elav* locus²⁰ (Fig. 3). In *sev* discs the R7 precursor does not express the *elav* antigen, which is indicative of its failure to adopt a neuronal fate. It develops instead as a non-neuronal cone cell²¹. In contrast, in *sev; sE-raf^{tor4021}* discs the R7 precursor does express this neural antigen (Fig. 3h), demonstrating that Raf activation can rescue the R7 precursor from adopting the cone cell fate in the absence of *sev* function. The R3 and R4 precursors, in which this construct is also expressed, are unaffected by ectopic Raf activity. Similarly, they do not respond to ectopic Sev^{22,23} or Ras1²⁴ activity, presumably because of a prior cell-fate commitment²³. The remaining *sev*-expressing cells, the non-neuronal mystery cells and cone cell precursors, do respond to ectopic Raf activity: they adopt a neuronal fate, as evidenced by their expression of the *elav* antigen (Fig. 3g, i). The diversion of these cells into a neuronal developmental pathway, also seen in response to ectopic Sev or Ras1 activity^{22,24}, accounts for the additional photoreceptor cells in the adult eye.

The observation that expression of either activated Sev, Ras1 or Raf proteins in all *sev*-expressing cells produce identical

phenotypes, and in particular that activation of either Ras1 or Raf is sufficient to rescue the R7 precursor in the absence of *sev* function, suggests that all three proteins act in a common pathway controlling specification of the R7 cell fate. Further evidence for this hypothesis is provided by the observation that the nuclear-localized Sina protein, normally required for Sev-induced R7 cell specification⁸, is also required for R7 development in response to either Ras1 or Raf activation (ref. 24; Fig. 1l).

To test whether Raf might act upstream or downstream of Ras1 in this pathway, we introduced the *raf^{HM7}* mutation into transgenic flies carrying an activated Ras1 construct, *sevRas1^{V12}* (ref. 24). Reduced Raf activity markedly suppresses the *sevRas1^{V12}* rough-eye phenotype (Fig. 1e, f and m, n), demonstrating that signalling by Ras1 is sensitive to the level of Raf function. Conversely, reducing *Ras1* gene dosage, which impairs signalling by either a hypo- or hyperactive Sev kinase (ref. 4; T. Raabe and E.H., unpublished data), does not appear to modulate the strength of signalling by Raf^{tor4021} (data not shown). Thus, Ras1 concentration is a rate-limiting factor in signalling by Sev, but not in signalling by Raf. These observations are consistent with a model in which Raf acts downstream of both Ras1 and Sev in a common signal transduction pathway, one target of which may be the nuclear Sina protein (Fig. 4).

These results bear a striking similarity to interactions observed between the Ras and Raf proteins in mammalian cells: transformation of NIH3T3 cells by *ras* oncogenes is dependent upon Raf-1 function²⁵, whereas transformation by *raf* oncogenes does not require cellular *ras* function²⁶. Furthermore, *p21^{c-ras}* mediates hyperphosphorylation of Raf-1 in PC12 cells in response to activation of the *trk* receptor tyrosine kinase²⁷. Thus the Raf serine/threonine kinase acts, either directly or indirectly, as an effector of Ras function in both vertebrates and invertebrates. Genetic screens for mutations that suppress the rough-eye phenotype of *sE-raf^{tor4021}* transformants should now enable the components acting downstream of Raf in such signal transduction pathways to be identified. □

Received 27 July; accepted 9 October 1992.

- Basler, K. & Hafen, E. *Cell* **54**, 299–312 (1988).
- Tomlinson, A., Bowtell, D. D. L., Hafen, E. & Rubin, G. M. *Cell* **51**, 143–150 (1987).
- Krämer, H., Cagan, R. L. & Zipursky, S. L. *Nature* **352**, 207–212 (1991).
- Simon, M. A., Bowtell, D., Dodson, G. S., Lavery, T. R. & Rubin, G. M. *Cell* **67**, 701–716 (1991).
- Rogge, R. D., Karlovich, C. A. & Banerjee, U. *Cell* **64**, 39–48 (1991).
- Bonfini, L., Karlovich, C. A., Dasgupta, C. & Banerjee, U. *Science* **255**, 603–606 (1992).
- Gaul, U., Mardon, G. & Rubin, G. M. *Cell* **68**, 1007–1019 (1992).
- Carthew, R. W. & Rubin, G. M. *Cell* **63**, 561–577 (1990).
- Cantley, L. C. et al. *Cell* **64**, 281–302 (1991).
- Li, P., Wood, K., Mamon, H., Haser, W. & Roberts, T. *Cell* **64**, 479–482 (1991).
- Perrimon, N., Engstrom, L. & Mahowald, A. P. *Dev Biol.* **110**, 480–491 (1985).

12. Nishida, Y. *et al.* *EMBO J.* **7**, 775–781 (1988).
13. Ambrosio, L., Mahowald, A. P. & Perrimon, N. *Nature* **342**, 288–291 (1989).
14. Kramers, P. G. N., Schalet, A. P., Paradi, E. & Huizer-Hoogteyting, L. *Mut. Res.* **107**, 187–201 (1983).
15. Klingler, M., Erdélyi, M., Szabad, J. & Nüsslein-Volhard, C. *Nature* **335**, 275–277 (1988).
16. Sprenger, F., Stevens, L. M. & Nüsslein-Volhard, C. *Nature* **338**, 478–483 (1989).
17. Basler, K., Siegrist, P. & Hafen, E. *EMBO J.* **8**, 2381–2386 (1989).
18. Karch, F., Török, I. & Tissières, A. *J. molec. Biol.* **148**, 219–230 (1981).
19. Heidecker, G. *et al.* *Molec. cell. Biol.* **10**, 2503–2512 (1990).
20. Robinow, S. & White, K. *J. Neurobiol.* **22**, 443–461 (1991).
21. Tomlinson, A. & Ready, D. F. *Science* **231**, 400–402 (1986).
22. Basler, K., Christen, B. & Hafen, E. *Cell* **64**, 1069–1082 (1991).
23. Van Vactor, D. L. J., Cagan, R. L., Krämer, H. & Zipursky, S. L. *Cell* **67**, 1145–1155 (1991).
24. Fortini, M. E., Simon, M. A. & Rubin, G. M. *Nature* **355**, 559–561 (1992).
25. Kolch, W., Heidecker, G., Lloyd, P. & Rapp, U. R. *Nature* **349**, 426–428 (1991).
26. Smith, M. R., DeGudicibus, S. J. & Stacey, D. W. *Nature* **320**, 540–543 (1986).
27. Wood, K. W., Sarnecki, C., Roberts, Y. M. & Blenis, J. *Cell* **68**, 1041–1050 (1992).
28. Klemenz, R., Weber, U. & Gehring, W. J. *Nucleic Acids Res.* **15**, 3947–3959 (1987).
29. Sprenger, F. & Nüsslein-Volhard, C. *Cell* (in the press).

ACKNOWLEDGEMENTS. We thank G. Rubin, M. Simon and N. Perrimon for fly stocks; K. White for anti-elav antibody; U. Jauch and T. Bächli for assistance with scanning electron microscopy and confocal microscopy, respectively; K. Basler for critical reading of the manuscript; and C. Hugentobler for technical assistance.

SEC21 is a gene required for ER to Golgi protein transport that encodes a subunit of a yeast coatomer

Midori Hosobuchi*, Thomas Kreis†
& Randy Schekman*

* Department of Molecular and Cell Biology and Howard Hughes Medical Institute, 401 Barker Hall, University of California, Berkeley, California 94720, USA

† Département de Biologie Cellulaire, Université de Genève, 30 quai Ernest-Ansermet, CH-1211 Genève 4, Switzerland

NON-CLATHRIN coated vesicles have been implicated in early steps of intercompartmental transport^{1–4}. A distinct set of coat proteins are peripherally associated with the exterior of purified mammalian intra-Golgi transport vesicles⁵. The 'coatomer', a cytosolic complex containing a similar subunit composition to and sharing at least one subunit (β -COP) with the coat found on vesicles, has been postulated to be the precursor of this non-clathrin coat⁶. Here we describe the characterization of *SEC21*, an essential gene required for protein transport from the endoplasmic reticulum to the Golgi in the yeast *Saccharomyces cerevisiae*. The 105K product of this gene, Sec21p, participates in a cytosolic complex that we show to be a yeast homologue of the mammalian coatomer. These observations demonstrate that a non-clathrin coat protein plays an essential role in intercompartmental transport.

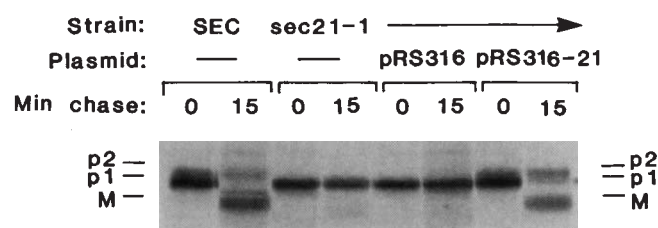
The *sec21-1* mutation was isolated in a selection for yeast cells that display a temperature-sensitive defect in secretion⁷,

and was determined to be deficient in protein transport from the endoplasmic reticulum (ER) to the Golgi^{8,9}. The *SEC21* gene was cloned from a yeast genomic library¹⁰ by complementation (M. Bernstein and R. S., unpublished results). The smallest insert was subcloned into the yeast integration vector pRS306 and the low-copy yeast shuttle vector pRS316 (ref. 11) creating the plasmids pRS306-21 and pRS316-21, respectively. The integrating plasmid pRS306-21 was used to demonstrate that the cloned DNA corresponds to the *sec21-1* locus. Briefly, a *MAT α SEC21 ura3* strain (RSY255) was transformed with pRS306-21 cut with *Bgl*III (which cuts the DNA within the insert) to direct integration of the plasmid marker, *URA3*, to the locus of the cloned DNA. Eight stable *Ura*⁺ integrants were isolated and mated to a *MAT α sec21-1 ura3* strain (RSY278). The resulting diploids were sporulated and 4–6 tetrads from each integrant cross were dissected, as well as 25 additional tetrads from one of the integrant crosses. In all tetrads analysed, *Ura*⁺*Ts*⁺ and *Ura*[–]*Ts*[–] phenotypes displayed absolute linkage (only parental ditypes were observed), indicating that the *URA3* marked locus of the cloned DNA is identical to the *sec21-1* locus.

The plasmid pRS316-21 complemented the temperature-sensitive secretion defect of *sec21-1* (Fig. 1), as determined by a pulse-chase *in vivo* labelling of wild-type (*SEC*) and *sec21-1* cells followed by immunoprecipitation of carboxypeptidase Y (CPY), an enzyme that travels through the secretory pathway *en route* to the vacuole¹². Wild-type cells grown at 37 °C (the nonpermissive temperature of *sec21-1*; Fig. 1, lanes 1 and 2) converted core-glycosylated CPY (p1) to Golgi (p2) and mature (M) forms in 15 min. The *sec21-1* mutant accumulated CPY in the p1 form when preshifted to 37 °C (Fig. 1, lanes 3 and 4). The vector pRS316 alone did not rescue this defect (Fig. 1, lanes 5 and 6), but pRS316-21, which contains the complementing insert, restored transport from the ER through the Golgi to the

FIG. 1 Immunoprecipitation of labelled carboxypeptidase Y in wild-type cells (*SEC*), *sec21-1* cells, and *sec21-1* cells containing plasmids without (pRS316) and with (pRS316-21) the *SEC21* gene insert. P1 CPY is the core-glycosylated form found in the ER, p2 is the outer chain-glycosylated form in the Golgi, and the mature form (M) results from proteolytic processing in the vacuole.

METHODS. Subcloning of *SEC21* gene: Six related plasmids were obtained from a multicopy YEp13 yeast genomic library constructed by Nasmyth¹⁰ by complementation of the temperature-sensitive growth defect of the *sec21-1* mutation (M. Bernstein and R. S., unpublished data). The smallest insert was subcloned into the low-copy yeast shuttle vector pRS316 (ref. 11) creating plasmid pRS316-21. Radiolabelling of cells: Cells were grown at 24 °C overnight in selective minimal medium containing 200 μ M ammonium sulphate²³ to an OD₆₀₀ of 0.25–0.5. After a 1-h shift to 37 °C the cells were transferred to sulphate-free selective minimal medium at 5 OD₆₀₀ ml^{–1}. After a preincubation for 5 min at 37 °C, 30 μ Ci per OD₆₀₀ unit of trans ³⁵S-label (ICN) were added to each tube. After 5 min, labelling was terminated by addition of 100 $\times$ chase cocktail (0.3% cysteine, 0.4% methionine and 100 mM ammonium sulphate), and 1.5 OD₆₀₀ chase time-points were removed as indicated and an equal volume of ice-cold 20 mM Na₂Na₃ was added. Cells were collected and washed with 10 mM Na₂Na₃. Glass beads (300 μ l) and 300 μ l lysis buffer (50 mM Tris-Cl pH 7.5, 1% SDS, 1 mM EDTA, 1 mM PMSF) were added, the mixture was agitated vigorously for 90 s at room temperature and then immersed in a boiling waterbath for 6 min. Immunoprecipitation, electrophoresis and autoradiography were done essentially as described¹⁴, using 3 μ l anti-CPY serum¹² and omitting the competing unlabelled cell extracts and the second round of precipitation.



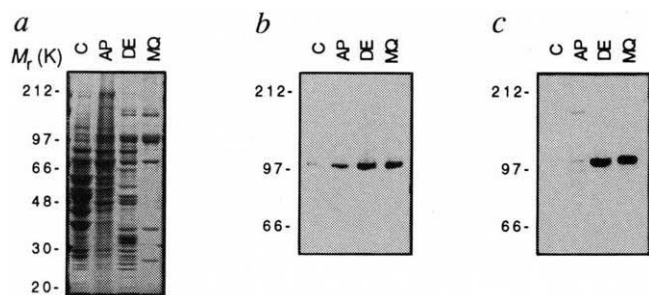


FIG. 2 Purification of Sec21p-containing complex from wild-type yeast cytosol. *a*, Gradient (3–17%) SDS-PAGE (Jule, New Haven, CT) followed by Coomassie stain analysis of TCA-precipitated⁶ fractions from the Sec21p purification. Lane 1, Yeast cytosol (C, 50 μ g), lane 2, ammonium sulphate pellet (AP, 25 μ g), lane 3, DEAE pool (DE, 25 μ g), lane 4 Mono Q pool (MQ, 2 μ g). *b*, Immunoblot of above fractions using affinity-purified antibodies against Sec21p. *c*, Immunoblot of above fractions using the 110-2 antibody against mammalian β -COP.

METHODS. Purification of Sec21p complex from yeast cytosol: Wild-type RSY607 yeast cells were grown to an OD₆₀₀ of 10 in YPD at 30 °C in a New Brunswick fermenter and collected in a Sharples continuous flow centrifuge. The cell pellet was washed with 0.15 M Buffer 88 (B88 20 mM K-HEPES, 250 mM sorbitol, 5 mM (CH₃COO)₂Mg and 0.15 M CH₃COOK) the molarity of B88 refers to concentration of acetate. The cells were resuspended in a minimal volume of 0.15 M B88, rapidly frozen in liquid nitrogen and lysed in a Waring blender²⁴. Lysed cells (300–400 g) were thawed to 4 °C (the rest of the preparation was completed at 0–4 °C), and diluted with 0.15 M B88 to 2 ml g⁻¹ cell weight. Protease inhibitors and DTT were added to the following concentrations: 0.5 mM PMSF (hereafter noted as P), 1 μ M leupeptin, 1 μ M pepstatin A, and 1 mM DTT (hereafter noted as D). The crude lysate was then centrifuged 10 min at 4,500g, 20 min at 26,000g, and 60 min at 100,000g. The protein concentration was determined by the method of Bradford²⁵ using BSA as a standard, and adjusted to 15–20 mg ml⁻¹ protein by the addition of B88PD. This fraction is yeast cytosol. Pulverized ammonium sulphate was added to 20% saturation, and the solution allowed to precipitate over a period of 30 min. The pellet was removed by centrifugation for

20 min at 26,000g and the supernatant was adjusted to 40% saturated ammonium sulphate, precipitated, centrifuged, and the resulting pellet was reserved. After rinsing the tubes briefly with fresh B88PD to remove adherent salt, the pellets were resuspended in 25–30 ml B88PD and the conductivity of the solution was adjusted to that of 0.5 M B88PD + 10% (w/v) glycerol (hereafter noted as G). The solution was cleared at 100,000g for 1 h to remove insoluble material, and was loaded on a 20 ml DEAE-Sephacose FF (Pharmacia) column equilibrated in 0.5 M B88GPD using a Pharmacia FPLC. Bound material was washed with 0.5 M and 0.625 M steps in B88GPD, and proteins eluted with a 175 ml 0.625–0.75 M linear gradient in B88GPD, followed by a 50 ml 0.75–1.0 M linear gradient in B88GPD. Fractions containing Sec21p immunoreactivity (elution at about 0.73 M K-acetate) were pooled and applied to a 1 ml Mono Q column (Pharmacia) equilibrated in 0.5 M B88GPD. The column was washed with 0.8 M B88GPD, and proteins eluted with a 64 ml linear gradient in B88GPD from 0.8 M–1.2 M. Fractions containing Sec21p immunoreactivity (elution at about 1.0 M potassium acetate) were pooled, dialysed against 0.15 M B88GPD and frozen. Typical preparations yielded about 1 mg of pure complex per 1 g yeast cytosol protein (roughly 10% yield of Sec21p). Preparation of affinity-purified antibodies: A Sec21p-protein A fusion was constructed using the cloned *SEC21* gene and was expressed in *Escherichia coli* using the pRIT vector system and host²⁶ and purified as described previously²⁷. Fusion protein (100 μ g) was injected into rabbits using the Freund's adjuvant system²⁸. Serum from rabbits containing Sec21p-specific antibodies was affinity purified against a bacterially expressed Sec21p-glutathione S transferase fusion protein (obtained using the pGEX-3X vector from Pharmacia LKB Biotechnology, Piscataway, NJ, and purified according to manufacturer's instructions) coupled to cyanogen bromide-activated (CNBr) Sepharose (Pharmacia LKB Biotechnology). Clarified sera were applied to the column and specifically-bound antibodies were eluted with 100 mM glycine pH 2.2 and stored in PBS²⁸. Affinity-purified Sec21 antibodies recognize a 105K protein that is overproduced in cells carrying *Sec21* on a multicopy plasmid (data not shown). The 110-2 antibody was produced by immunizing rabbits with the 110-2 peptide, (K)IHADQDRALDYLT (single-letter amino-acid code), coupled to keyhole limpet haemocyanin, followed by affinity purification using the peptide coupled to CNBr Sepharose as described previously²⁹. Both antibodies were used at a dilution of 1/200. ¹²⁵I-protein A immunoblots³⁰ were done as described previously, except the ¹²⁵I-protein A concentration was as described in ref. 6. Blots were visualized using a Phosphorimager (Molecular Dynamics, Sunnyvale, CA).

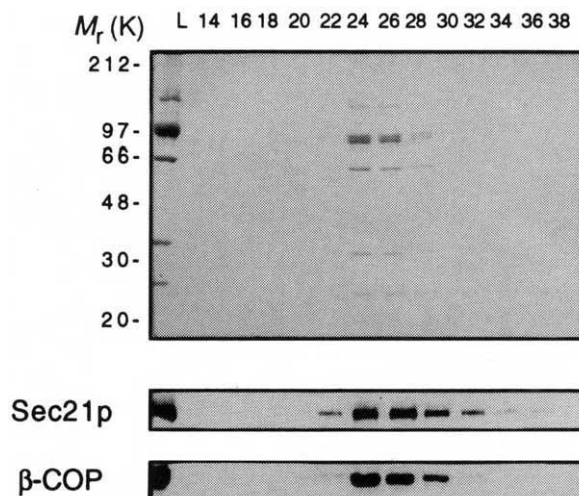
vacuole with near wild-type kinetics (Fig. 1, lanes 7 and 8). These data demonstrate that the *SEC21* gene is required for ER to Golgi protein transport.

The DNA sequence of the minimal complementing fragment was determined by the Sanger method¹³ (GenBank accession number M59708). The predicted protein sequence of the *SEC21* gene product indicated a hydrophilic protein with a relative molecular mass of 105K. Deletion of this putative coding region was lethal as determined by tetrad analysis (data not shown), suggesting that the function encoded by *SEC21* is also essential for growth.

Subcellular fractionation of yeast extracts by differential centrifugation followed by immunoblotting using a Sec21p-specific affinity-purified antibody (see Fig. 2 legend) showed that Sec21p is present in both a soluble pool and a sedimentable fraction¹⁴. The Sec21p present in the sedimentable fraction was extracted with urea and sodium carbonate, pH 11.5, as well as detergents such as Triton X-100 (ref. 14), suggesting that the sedimentable form of Sec21p is peripherally associated with membranes. The soluble form of Sec21p in unfractionated cytosol exists in a complex of about 700–800K as determined by gel filtration chromatography (data not shown), suggesting that Sec21p is

FIG. 3 Analysis of purified Sec21p-containing complex. Gel filtration chromatography of Mono Q dialysed pool. Top panel, 3–17% gradient SDS-PAGE followed by Coomassie stain analysis of fractions obtained from gel filtration. Middle and bottom panels, immunoblots of above fractions using affinity-purified antibodies against Sec21p or β -COP.

METHODS. Frozen purified Sec21p complex was cleared at 100,000g for 15 min and a 200 μ l aliquot (100 μ g) was applied to a 21 ml Superose-6 column (Pharmacia) equilibrated in 0.15 M B88GPD. Fractions (0.5 ml) were collected and 40 μ l of each fraction and 4 μ l of the load fraction were TCA-precipitated and analysed as described⁶. Column calibration standards were obtained from Pharmacia and used according to manufacturer's instructions. Standards fractionated as follows: void, fraction 15; thyroglobulin (669K), fraction 26; ferritin (440K), fraction 29; catalase (232K), fraction 31; BSA (68K), fraction 33; CoCl₂ (238), fraction 42.



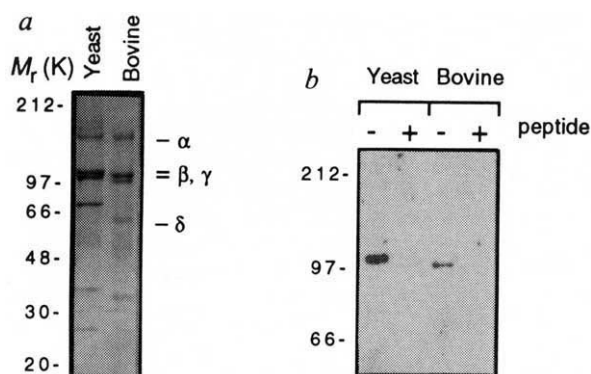


FIG. 4 Comparison of the Sec21p complex with bovine coatamer. *a*, Analysis of purified Sec21p complex and mammalian coatamer by 3–17% gradient SDS-PAGE followed by Coomassie stain analysis. *b*, Immunoblot analysis of purified Sec21p complex and mammalian coatamer using the 110–2 anti-β-COP antibody incubated in the absence (odd lanes, –) and presence (even lanes, +) of competing peptide.

METHODS. *a*, Purified Sec21p complex (3 μg) and coatamer obtained from bovine brain⁶ (3 μg; gift from M. Gerard Waters) were analysed by 3–17% SDS-PAGE followed by Coomassie staining. *b*, Purified Sec21p complex (1 μg) and bovine coatamer (0.1 μg) were transferred to nitrocellulose, stained with ponceau S and individual lanes were marked and cut into strips. Strips corresponding to even lanes were immunoblotted with 110–2 antibody as described above. Strips corresponding to odd lanes were immunoblotted with 110–2 antibody preincubated with a 5,000 molar excess of 110–2 peptide (see legend to Fig. 2) for 15 min at room temperature before diluting 1/200 in 2% milk buffer. Before exposure, blots were reconstructed using markings as a reference.

contained in a protein complex with itself or other proteins. To understand the nature of this complex, Sec21p was purified to homogeneity from yeast cytosol using immunoblotting as an assay (Fig. 2*a, b*). The three purification steps used to obtain pure Sec21p-containing complex from yeast cytosol (Fig. 2*a*, lane 1) were: ammonium sulphate precipitation (Fig. 2*a*, lane 2), DEAE Sepharose chromatography (Fig. 2*a*, lane 3), and Mono Q chromatography (Fig. 2*a*, lane 4). The resulting set of six major polypeptides (Fig. 2*a*, lane 4) exactly copurify with each other and with the Sec21p immunoreactivity over a gel filtration column (Fig. 3, top and middle panels), as well as in four additional steps: Mono S column and hydroxylapatite column chromatography, velocity sedimentation on a glycerol gradient, and a low salt/low pH precipitation step (data not shown).

The native M_r of the purified Sec21p-containing complex is 700–800K (Fig. 3, top panel). The subunit profile of the complex, with polypeptides of 150, 110, 105, 73, 35 and 25K (Fig. 4*a*, lane 1), closely resembled that of mammalian coatamer isolated from bovine brain (Fig. 4*a*, lane 2), a protein complex of 700–800K native M_r that has been postulated to be the coat protomer of intra-Golgi transport vesicles⁶. The other subunits of the complex do not correspond to any characterized SEC gene products. The same purification steps used to isolate the mammalian coatamer⁶ were used either to isolate the Sec21p-containing complex (Fig. 2*a*) or to verify its purity (see above). To confirm that the Sec21p-containing complex represented a yeast coatamer, we analysed fractions from the Sec21p purification by immunoblot using an antibody against the 110K protein β-COP, the best characterized component of the mammalian coatamer^{15–18}. The polyclonal anti-peptide 110–2 antibody was chosen because crossreaction was detected with a putative β-COP in *Xenopus* oocyte extracts. This antibody recognized the 110K protein (thus defined as the yeast β-subunit; Sec21p corresponds to the γ-subunit of the yeast complex) of the Sec21p-containing complex (Fig. 4*b*, lane 1). Addition of

the peptide to which the antibody was raised eliminated all crossreactivity with either the yeast (Fig. 4*b*, lane 2) or bovine (Fig. 4*b*, lane 4) subunits, indicating that the recognition is indeed specific. Furthermore, the pattern of enrichment of the β-COP crossreacting species across the purification was similar to that of Sec21p (Fig. 2*b, c*). Although other crossreacting species were detected in one of the yeast fractions (AP; Fig. 2*c*, lane 2), the ~95K crossreacting species is always present in minor amounts and the ~160K crossreacting species is not always present. Lastly, the β-COP crossreacting species exactly copurified with the yeast complex and Sec21p over the last purification step (Fig. 3*a*, all panels). The gene that encodes the yeast β-COP has been obtained and complete sequence analysis demonstrates a 43% identity with mammalian β-COP¹⁶ (R. Duden and R. S., unpublished results). The presence of this β-COP homologue, together with the similar subunit composition, size and purification properties, indicates that this Sec21p-containing complex of yeast is a homologue of the mammalian coatamer.

Previous studies have indicated a circumstantial connection between the appearance of non-clathrin coated vesicles and protein transport between cisternae of the mammalian Golgi complex^{1–4}. We have characterized the SEC21 gene, which encodes a subunit of a yeast non-clathrin coatamer, and have shown for the first time that such a protein is required for intercompartmental transport. Because SEC21 function is required for ER to Golgi transport and, in contrast to clathrin heavy¹⁹ and light chains²⁰, is essential for yeast growth, it will be possible to use genetic methods to examine further the function of the Sec21p-containing coatamer *in vivo*. In addition, the availability of a yeast *in vitro* ER to Golgi transport assay²¹ capable of separating the steps of budding from the ER, targeting to the Golgi, and fusion with the *cis*-Golgi²² will allow the determination of the precise role of the non-clathrin coat in intercompartmental transport. □

Received 8 October; accepted 4 November 1992.

- Orci, L., Glick, B. S. & Rothman, J. E. *Cell* **46**, 171–184 (1986).
- Melançon, P. et al. *Cell* **51**, 1053–1062 (1987).
- Orci, L., Malhotra, V., Amherdt, M., Serafini, T. & Rothman, J. E. *Cell* **56**, 357–368 (1989).
- Malhotra, V., Serafini, T., Orci, L., Shepherd, J. C. & Rothman, J. E. *Cell* **58**, 329–336 (1989).
- Serafini, T. et al. *Nature* **349**, 215–220 (1991).
- Waters, M. G., Serafini, T. & Rothman, J. E. *Nature* **349**, 248–251 (1991).
- Novick, P., Field, C. & Schekman, R. *Cell* **21**, 205–215 (1980).
- Novick, P., Ferro, S. & Schekman, R. *Cell* **25**, 461–469 (1981).
- Kaiser, C. & Schekman, R. *Cell* **61**, 723–733 (1990).
- Nasmyth, K. A. & Reed, S. I. *Proc. natn. Acad. Sci. U.S.A.* **77**, 2119–2123 (1980).
- Sikorski, R. S. & Hieter, P. *Genetics* **122**, 19–27 (1989).
- Stevens, T., Esmon, B. & Schekman, R. *Cell* **30**, 439–448 (1982).
- Sanger, F. S., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467 (1977).
- Stirling, C. J., Rothblatt, J. A., Hosobuchi, M., Deshaies, R. & Schekman, R. *Molec. Biol. Cell* **3**, 129–142 (1992).
- Duden, R., Allan, V. & Kreis, T. *Trends Cell Biol.* **1**, 14–19 (1991).
- Duden, R., Griffiths, G., Frank, R., Argos, P. & Kreis, T. E. *Cell* **64**, 649–665 (1991).
- Klausner, R. D., Donaldson, J. G. & Lippincott-Schwartz, J. *J. Cell Biol.* **116**, 1071–1080 (1992).
- Rothman, J. E. & Orci, L. *Nature* **355**, 409–415 (1992).

- Payne, G. S., Hasson, T. B., Hasson, M. S. & Schekman, R. *Molec. cell. Biol.* **7**, 3888–3898 (1987).
- Silveira, L. A., Wong, D. H., Maslarsz, F. R. & Schekman, R. *J. Cell Biol.* **111**, 1437–1449 (1990).
- Baker, D., Hicke, L., Rexach, M. & Schekman, R. *Cell* **54**, 335–344 (1988).
- Rexach, M. F. & Schekman, R. *J. Cell Biol.* **114**, 219–229 (1991).
- Sherman, F. in *Guide to Yeast Genetics and Molecular Biology* (eds Guthrie, C. & Fink, G. R.) 3–21 (Academic, San Diego, 1991).
- Dunn, B. & Wobbe, C. R. in *Current Protocols in Molecular Biology* Vol. 2 (eds Ausubel, F. M. et al.) 13.13.5–13.13.6 (Greene/Wiley Interscience, New York, 1989).
- Bradford, M. *Analyt. Biochem.* **72**, 248–250 (1976).
- Stanley, K. K. & Luzio, J. P. *EMBO J.* **3**, 1429–1434 (1984).
- Feldheim, D. A., Rothblatt, J. A. & Schekman, R. *Molec. Cell Biol.* **12**, 3288–3296 (1992).
- Harlow, E. & Lane, D. *Antibodies—A Laboratory Manual* **98**, 100–105, 114 (Cold Spring Harbor Laboratory Press, New York, 1988).
- Kreis, T. E. *EMBO J.* **5**, 931–941 (1986).
- Hicke, L. & Schekman, R. *EMBO J.* **8**, 1677–1684 (1989).

ACKNOWLEDGEMENTS. We thank M. G. Waters for providing purified bovine coatamer, C. Barlowe, J. Brodsky, R. Duden, M. Rexach and A. Sidow for comments on the manuscript, and members of the Schekman laboratory for scientific discussion. The sequence described here has been submitted to GenBank, accession number M59708.

A whole genome approach to *in vivo* DNA-protein interactions in *E. coli*

Ming X. Wang* & George M. Church†

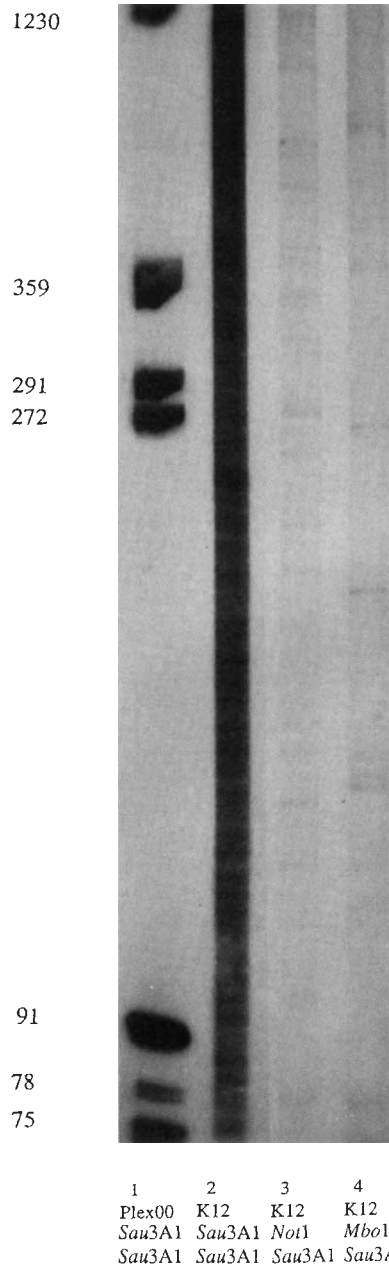
* Laboratory of Oncology Research, Research Division, Wills Eye Hospital and Jefferson Medical College of Thomas Jefferson University, Philadelphia, Pennsylvania 19107, USA
 † Department of Genetics, Harvard Medical School and Howard Hughes Medical Institute, Boston, Massachusetts 02115, USA

THE increasingly rapid pace at which genomic DNA sequences are being determined has created a need for more efficient techniques to determine which parts of these sequences are bound *in vivo* by the proteins controlling processes such as gene expression, DNA replication and chromosomal mechanics. Here we describe a whole-genome approach to identify and characterize such DNA sequences. The method uses endogenous or artificially introduced methylases to methylate all genomic targets except those protected *in vivo* by protein or non-protein factors interfering with methylase action. These protected targets remain unmethylated in purified genomic DNA and are identified using methylation-sensitive restriction endonucleases. When the method was applied to the *Escherichia coli* genome, 0.1% of the endogenous adenine methyltransferase (Dam methylase) targets were found to be unmethylated. Five foreign methylases were examined by transfection. Database-matched DNA sequences flanking the *in vivo*-protected Dam sites all fell in the non-coding regions of seven *E. coli* operons (*mtl*, *cdd*, *flh*, *gut*, *car*, *psp* and *fep*). In the first four operons these DNA sequences closely matched the consensus sequence that

binds to the cyclic AMP-receptor protein. The *in vivo* protection at the Dam site upstream of the *car* operon was correlated with a downregulation of *car* expression, as expected of a feedback repressor-binding model.

Our method is based on three assays: (1) characterizing the genomic DNA patterns of the *in vivo*-protected methylation targets using end-labelling and gel fractionation; (2) cloning and sequencing of these sites to identify both exact DNA matches and more degenerate protein-binding motifs in databases; (3) direct physiological tests to determine partial *in vivo* protection of the target as a function of genetic and environmental changes by means of filter hybridization assays. The *E. coli* genome provides a particularly good testing system for our method as over 40% of the genome has been sequenced. We examined the genomic DNA sequence GATC by using the endogenous Dam methylase which methylates the N6 position of adenines in GATC sequences. Figure 1 shows an end-labelling experiment which demonstrates that about 20 GATC sites are completely unmethylated (0.1% of the total targets in the genome) and that many sites are partially methylated¹⁻⁶. Methylation targets remain unmethylated as a result of binding of

FIG. 1 Autoradiograph of an end-labelling experiment. Lane 1 is a Plex00 molecular weight standard (sizes given in nucleotides); other lanes represent *E. coli* genomic DNA cut by different combinations of endonucleases as follows: 5 µg genomic DNA was digested with restriction endonuclease I, end-labelled with 1 µl reverse transcriptase (10 U µl⁻¹; Stratagene) and 0.5 µl [α -³²P]dNTP (1 mCi in 50 µl), chosen to correspond to the base following the restriction site 3' terminus, which helped to reduce background labelling at random breaks. After incubation at 37 °C for 40 min, the reaction was chased with 1 µl of each dNTP at 100 mM, digested with 2 µl endonuclease II (chosen to cut the genome frequently), then electrophoresed on a 6% non-denaturing polyacrylamide gel. The gel was dried and autoradiographed. Restriction endonucleases *Mbo*I, *Sau*3A1 and *Not*I were used. *Mbo*I specifically cleaves only GATC sequences with unmethylated adenines; *Sau*3A1 acts as a control, cleaving at this sequence regardless of its adenine methylation status (about 21,000 times in the *E. coli* genome based on current sequence data); *Not*I is a single-copy intensity control which we used to help estimate the fraction of bands in the *Mbo*I digest that are especially intense owing to comigration of multiple bands or are noticeably weak as a result of partial site protection. Restriction endonucleases I and II are used in different lanes as follows: lane 2: *Sau*3A1/*Sau*3A1, representing all genomic DNA fragments containing a GATC sequence at each end; lane 3: *Not*I/*Sau*3A1, indicating single-copy intensity control; lane 4: *Mbo*I/*Sau*3A1, representing genomic fragments bearing an unmethylated GATC (*Mbo*I) site at one end and a *Sau*3A1 site at the other. About 20 prominent bands corresponding to completely unmethylated GATC sites are seen; many partially methylated sites (faint bands) are also evident. Unmethylated GATC sites were cloned as follows: 100 µg *E. coli* genomic DNA was cut with *Mbo*I, extracted twice with phenol, mixed with 1 µg *Bam*HI-cut Bluescript SK vector (Stratagene), precipitated with ethanol and resuspended in 50 µl H₂O; 2 µl each of ATP (100 mM) and T4 DNA ligase (2,000 U per µl) were then added. The ligation mixture was size-fractionated on a 1% agarose gel and any products larger than the major vector bands were excised and purified using GeneClean (Bio101), then resuspended in 50 µl H₂O. A *Cla*I reaction liberated small (6 kb on average) vector-genomic ligation products, and after phenol extraction and ethanol precipitation, a ligation reaction was done to circularize the products. *E. coli* cells (BRL Max Efficiency DH5 α) were then transformed to produce the libraries, from which random plasmid clones were purified and 5 µl of each plasmid was treated with 0.5 µl RNaseA (10 µg µl⁻¹), denatured with 1 µl NaOH (2 M), precipitated with ethanol and sequenced using Sequenase (USB).



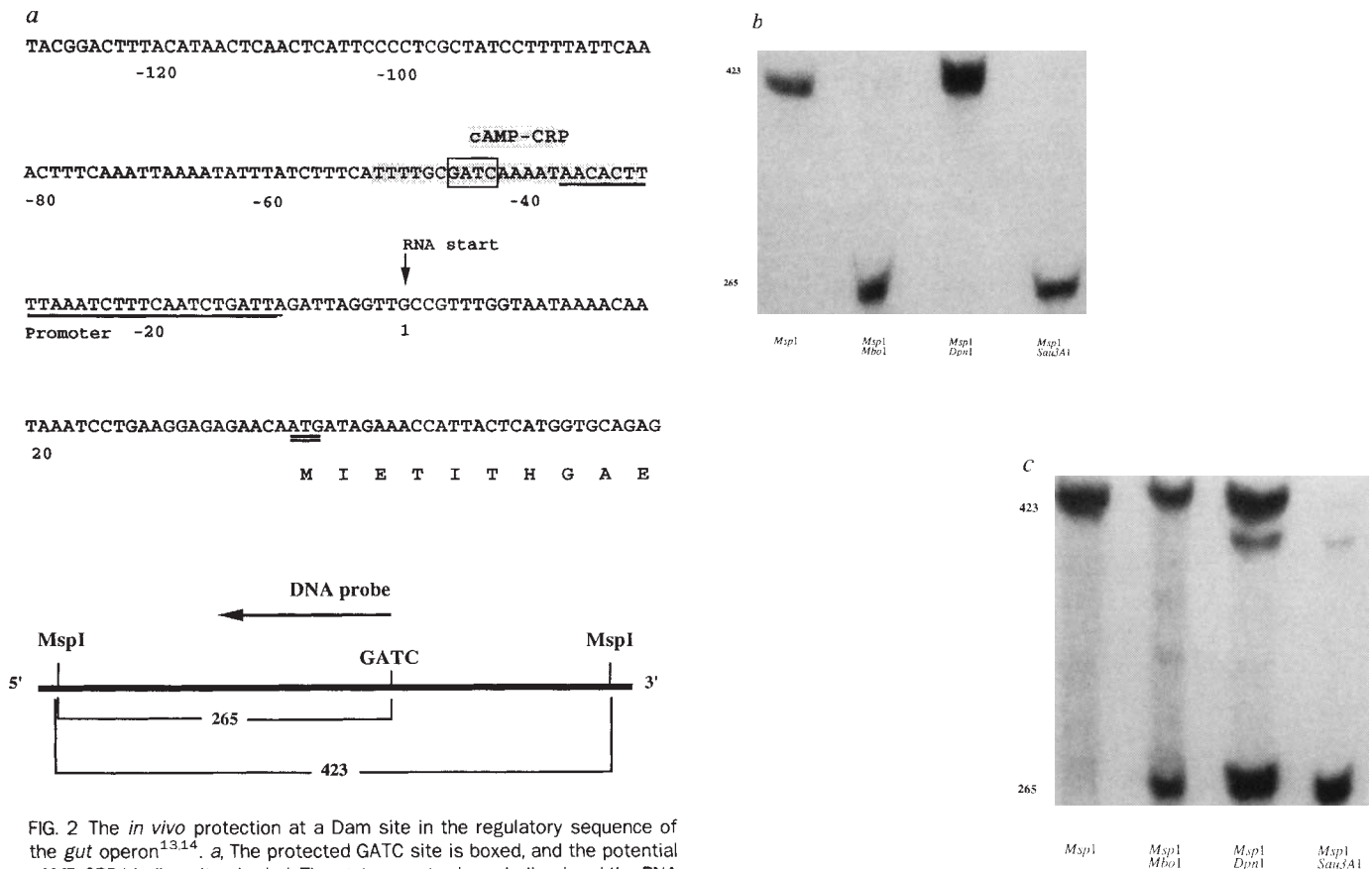


FIG. 2 The *in vivo* protection at a *Dam* site in the regulatory sequence of the *gut* operon^{13,14}. **a**, The protected GATC site is boxed, and the potential cAMP-CRP binding site shaded. The *gut* promoter is underlined and the RNA start site is indicated with an arrow; the translational start site ATG is underlined twice. Nucleotide numbering starts with the RNA start site. The diagram illustrates the relationship between the *in vivo*-protected GATC site, the flanking restriction endonuclease sites, and the DNA probe which are used for filter hybridization (**b** and **c**). **b**, Autoradiograph of a filter hybridization experiment using a DNA probe complementary to the regulatory region of the *gut* operon. DNA was extracted from *E. coli* K-12 EMG2 grown in rich medium (1% yeast extract, 2% tryptone) to early stationary phase. After endonuclease digestion, the samples were electrophoresed on 6% denaturing polyacrylamide gels and electrotransferred to nylon, crosslinked by ultraviolet irradiation and hybridized^{37,38}. The restriction enzymes used for each sample are indicated under the lanes, see **a** for explanation of the hybridization bands. Numbers indicate sizes in nucleotides. The *gut* DNA probe was constructed by 3' extension from a 20-nucleotide primer annealed to the vector near the cloning site: 10 pmol of the Bluescript plasmid (Stratgene) containing the *gut* inserts were treated with 2 μ l RNase A (10 U μ l⁻¹), denatured with 4 μ l 2M NaOH, precipitated with ethanol and

air-dried. It was then incubated with annealing mix (2 μ mol 20-nucleotide primer per 4 μ l USB 10 $\times$ Sequenase buffer and 12 μ l water) at 37 $^{\circ}$ C for 30 min. The probe was labelled with 4 μ l (20 pmol) of [α -³²P]dATP, 1 nmol each of dCTP, dGTP and dTTP using 2.5 μ l (32 U) of USB Sequenase. The reaction products were denatured in formamide at 90 $^{\circ}$ C and fractionated on a 6% acrylamide-urea gel. A narrow region of the gel (80–120 nucleotides) was excised and eluted by grinding in hybridization buffer (7% SDS, 10% polyethylene glycol, 0.25 M NaCl)^{37,38}. **c**, Autoradiograph of filter hybridization experiments using an *E. coli* strain with a deletion in the *crp* gene which inactivates the CRP¹⁵. *E. coli* K12 CGSC 7043 (λ -*relA1 spoT1 thi1 rpsL136* Δ *crp45*) cells were grown in minimal A medium with 1% glycerol to early stationary phase. DNA was prepared and hybridized as **b**. The additional weaker bands in lanes 3 and 4 were due to low-stringency hybridization and were not normally seen. Further comparison using two *E. coli* strains (isogenic except the *crp* locus) show 82% *in vivo* protection in the *crp*⁺ strain and 6% protection in the *crp*⁻ strain.

protein or non-protein factors, DNA conformational steric hindrance⁷ or demethylation by DNA repair. These we refer to collectively as protecting factors. A typical *E. coli* DNA-binding protein sterically blocks 10 to 70 base pairs of target DNA from access to DNase I (refs 8, 9). Hence each GATC site could lie within footprints anywhere in a region of 22 to 142 base pairs and the sum of these footprints covers about 10–60% of the *E. coli* genome. To expand the genomic portion that could be assayed, we studied foreign methylases introduced by transfection. Prokaryotic methylases, with 111 types characterized and cloned¹⁰, make a rich source of target sequences experimentally accessible. Perturbing effects of methylases on cellular processes might confine experiments to lower intracellular methylase concentrations or to brief methylase induction periods. We explored the possible use of exogenous methylases by examining five *E. coli* strains transfected with one of the following foreign methylase genes each: *HhaI*, *HhaII*, *HpaII*, *MspI* and *TaqI*, having corresponding target DNA sequences: GCGC, GANTC, CCGG, CCGG and TCGA respectively. Methylase overproduction resulted in no obvious phenotypic effect, and the

end-labelling experiments produced bands similar in number but distinct in pattern from that of *Dam*.

The cloning strategy (Fig. 1 legend) eliminated the initially high levels of background clones, such that essentially all genomic DNA fragments cloned contained an unmethylated GATC site each. The flanking DNA sequences of seven of the first nine such GATC sites cloned matched to the *E. coli* database in seven different operons: *mtl*, *cdd*, *flh*, *gut*, *car*, *psp* and *fep*. Remarkably, all seven *in vivo*-protected GATC sites fall in the 5' non-coding regions of these genes. As the probability that this will occur by chance is extremely small (1 in 10⁶ on the basis of current *E. coli* sequence data), our finding indicates that the protecting factors have a strong and non-random preference for the upstream regulatory regions of *E. coli* genes. Table 1 lists these DNA sequences and shows that the *in vivo*-protected GATC sites are located between -230 and +5 base pairs relative to the transcription start sites. DNA sequences flanking these GATCs in four operons (*mtl*, *cdd*, *flh* and *gut*) show close matches to the consensus sequence that binds to cAMP-receptor protein (CRP). In the *cdd* operon, the putative CRP-binding

sequence, previously shown to bind to CRP strongly *in vitro*¹¹, is found here to contain an *in vivo*-protected GATC site. In the *mtl* operon, the protected GATC locus overlaps the most significant CRP consensus match in that region. In contrast to the above four operons, the regulatory sequences of the *car*, *psp* and *fep* operons, which were not thought to be under CRP control, lack significant matches to the CRP box. The protection at the GATC loci in the regulatory regions of these operons may be due to interactions by other factors that, for example, regulate carbamoyl phosphate synthesis (*car*), stress response (*psp*), and iron transport (*fep*). The fraction of each Dam site

protected in a cell population was quantitated by filter hybridization, a method similar to those used to study vertebrate CG methylation¹². The genomic DNA from cell cultures was analysed by appropriate methylation-sensitive restriction endonucleases and specific DNA probes were used to visualize selectively the *in vivo* protection of each methylation target in the genomes directly and without cloning. These results are also shown in Table 1.

The *gut* operon is responsible for glucitol uptake in *E. coli*^{13,14}. The undermethylated GATC locus in the *gut* regulatory region (Fig. 2a) shows the strongest *in vivo* protection (95%; Fig. 2b),

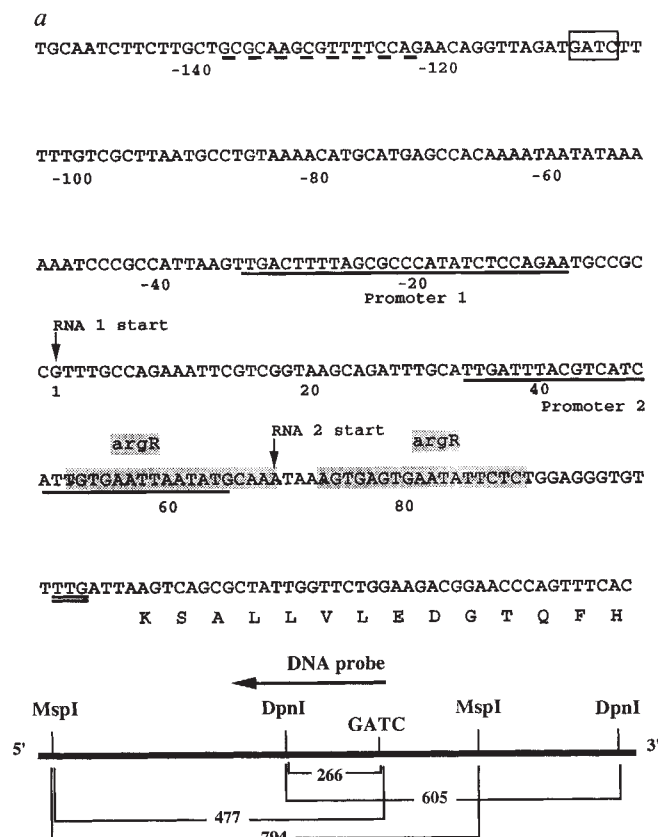


FIG. 3 The *in vivo* protection at a Dam site in the regulatory region of the *car* operon. **a**, The protected GATC site is boxed. A *purR* box is shown with a dashed underline. Promoters P1 and P2 are underlined, and the corresponding RNA start sites are indicated by arrows. Two *argR* boxes are represented by tandem shaded rectangles. The translational start site for *carA* initiation is underlined twice. Nucleotide numbering starts with the RNA1 start site. The diagram shows the relationship between the undermethylated GATC site, the flanking restriction endonuclease sites and the DNA probe which were used in the hybridization reactions (**b** and **c**). **b**, Autoradiograph of filter hybridization reaction using a DNA probe complementary to the *car* regulatory sequence. Strains and media were as described in Fig. 2b; hybridization bands are explained in **a**. Numbers indicate sizes in nucleotides. The DNA probe was constructed from a plasmid clone containing the appropriate *car* insert using methods described for Fig. 2b. Restriction enzymes used are indicated under the appropriate lanes. **c**, Hybridization band patterns arising from *E. coli* EMG2 cells grown in nutrient conditions with different availability of pyrimidine and arginine. Lane 1, minimal A medium; lane 2, minimal A plus uracil (500 $\mu\text{g ml}^{-1}$) and cytidine (1 mg ml^{-1}); lane 3, minimal A plus L-arginine (1 mg ml^{-1}); lane 4, minimal A plus uracil, cytidine and arginine. The degrees of *in vivo* protection at this *car* GATC locus (see **a**) are 0.05%, 3%, 0.07% and 15%, respectively. Hence, without pyrimidines the GATC site is essentially unprotected and independent of arginine. Pyrimidines have a significant effect and seem to be synergistic with arginine. The finding that arginine availability by itself does not affect the methylation status of the GATC site is consistent with the previous assignment¹⁶ of arginine repressor interaction at the P2 promoter, which is located at quite a distance (138 bp) downstream from the GATC site that is pyrimidine-responsive.

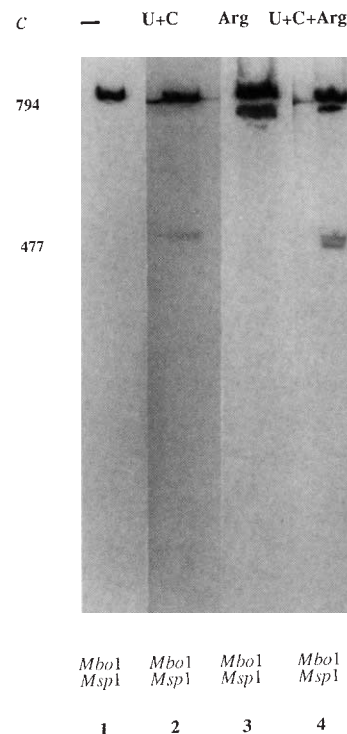
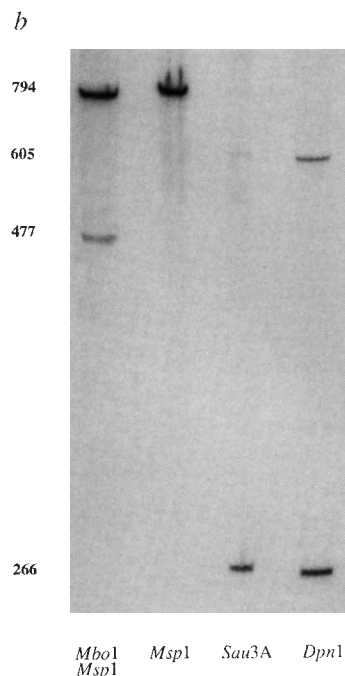


TABLE 1 DNA sequences flanking *in vivo*-protected GATC sites in the regulatory regions of seven *E. coli* operons and comparison with the consensus CRP-binding sequence

Gene	Map	Unmethylated (%)		Position	CRP consensus matches		Match score (s.d. units)
		<i>MboI</i>	<i>DpnI</i>				
<i>mtl</i>	81	—	—	—261	5' TAACAT G CTGT	AGATC A CATCA	2.7
		14	20	—218	5' TCTTGT GATTC	AGATC A CAAAAT	6.2
		—	—	—180	5' AAATGT GACAC	TACT C ACATTT	6.6
		—	—	—107	5' TTTTGT GATGA	ACG T CACGTCA	5.0
<i>cdd</i>	46	—	—	—63	5' TTATGT GATTG	AT A TCACAA	5.5
		—	—	—97	5' ATTTG C GATGC	GTC G CGATTTT	0.6
		18	22	—41	5' TAATG A GATTC	AGATC A CATAT	4.4
<i>flh</i>	42	38	46	—230	5' ----- GATC T	GTCATC A CGAA	—0.3 to 4.2
		—	—	—13	5' TGC G GT G AAAC	CGCTAA A ATA	0.2
<i>gut</i>	58	89	99	—47	5' TTTT G C GATCA	AA A TAA C ACTT	1.2
<i>car</i>	1	14	21	—107	5' TTAGAT GATCT	TTT T GT C GCTT	—1.9
<i>psp</i>	29	ND	—	—16	5' ATTCTT C AATC	AGAT C TTTATA	—2.2
<i>fep</i>	13	ND	—	+5	5' ATAT C CAATA	AGAT C GATAAC	—2.8
CRP consensus:					5' aaaTGT Gat ct	aga TC ACAtt t	8.5

The first column gives the name of each genetic locus, the next column the location on the genetic map in minutes. The third column shows the per cent unmethylation (protection) assayed by *MboI* and *DpnI* respectively through densitometry of filter hybridization bands. We selected *DpnI* whose specificity is in a sense complementary to *MboI* in that it recognizes only GATC sites methylated on both strands²⁶. In every case the *MboI* estimate is consistently lower by 4 to 10% than that of *DpnI*, which could be due to partial enzyme cleavage or to hemimethylation (as neither enzyme cleaves hemimethylated DNA²⁶). We determined the upper bounds to partial enzyme cleavage to be 3% based on the band intensities of adjacent GATC sites. The 'position' column lists the position of the G in GATC relative to the RNA start site (for segments lacking a GATC, it refers to the seventh base from the left of the CRP box). Bases matching the GATC sites in the CRP consensus are in bold. Undermethylated GATC sites are underlined. The rightmost column lists the score for a match to the consensus CRP-binding sequence. The scores are in standard deviation units (s.d.) above the mean for all *E. coli* sequences normalized to a length of 22 bp using the GCG version of the program ProfileSearch²⁷. A reasonable indicator of a significant match is +2 s.d. The range covers from -2.8 to +8.5 s.d. All CRP sites noted previously^{11,13,28,29-35} are included. Promoters for *flh* and *psp* lack S1 or reverse transcriptase mapping of the transcription start, instead putative -10 sequence motifs were used for alignment. A symmetric CRP consensus matrix was built by including both orientations of known CRP sites^{11,13,28,29-35}, which excluded all sites analysed here. In the CRP consensus binding sequence, the most commonly found base is given and the upper-case letters represent the most highly conserved base pairs which have been modelled to be in contact with the symmetric CRP protein dimer^{9,36}. ND, not determined.

even though the CRP box spanning this Dam site has the lowest match score among the four CRP-regulated operons. To investigate the possible involvement of CRP, we examined this GATC locus in an *E. coli* strain with a deletion in the *crp* gene which inactivated CRP¹⁵ and found a significantly reduced level of protection (50%; Fig. 2c). The change could be due to strain or environmental variations, or to an indirect effect of CRP on other proteins or nearby DNA conformations, but probably represents direct CRP binding.

The *car* operon encodes carbamoyl phosphate synthetase, a common element of the arginine and pyrimidine biosynthetic pathways. Pyrimidines and arginine repress the transcription of *car*, although the locations of protein-binding sites were uncertain¹⁶. We found an *in vivo*-protected GATC locus in the *car* regulatory region (Fig. 3a and b; 18% protection). Evidence that DNA-protein binding around this locus regulates *car* expression was obtained by examining *in vivo* protection as a function of *E. coli* growth conditions which differed in pyrimidine and arginine availability. As shown in Fig. 3c, there is no protection at this *car* GATC locus without pyrimidines, 3% protection with pyrimidines alone, and 15% when both nutrients are present. Our finding indicates that *in vivo* protection at this *car* upstream regulatory sequences is probably due to the binding of pyrimidine repressor(s). The arginine repressor acts (*argR*; Fig. 3a) synergistically with pyrimidines in down-regulating *car* expression.

The degree of methylation protection can be correlated with that of protein-binding through kinetic modelling because protein factors need to bind to DNA targets persistently to prevent methylation, whereas methylases need only a brief contact with targets to succeed in methylation. If one assumes that the methylase acts on target sequences in a genome at random, then the fraction of targets methylated per unit time is constant and the kinetic interaction between methylases and protecting factors is characterized by Poisson first-order decay: $U = 2^{-[(1-P)G/T]}$, where P is the fraction of a cell generation with protein factors that bind to the target sequence, U the unmethylated fraction of the target site at steady state, T the half-life of methylase

action^{17,18}, and G the cell generation time. For the *in vivo*-protected *gut* GATC locus, $U = 0.95$, so $P = 0.99$, that is, protein factors bind to this DNA sequence for 99% of the cell cycle. For the GATC site in the *car* operon, $U = 0.18$, and hence $P = 78\%$.

Our method of studying *in vivo* DNA-protein interactions has several advantages over previous *in vitro* and *in vivo* footprinting techniques. *In vitro* experiments allow fractionation and modification of the interacting components but reflect artefactual deviations from intracellular states. *In vivo* footprinting techniques¹⁹⁻²² typically require chemistry hazardous to cells and hence may alter chromatin structure. Our enzymatic method is less perturbing to cell integrity. Furthermore, it avoids assumptions about protein-binding sites and so is applicable for genome-scale analysis. Such methods can be applied^{18,23} to the study of regulation of individual genes, as well as to overall changes in pattern in the genome and in chromosomal structures. Extension to other organisms and target sequences is dependent on the efficiency of expression of transfected methylase genes and on the compatibility of the methylation with cellular physiology. In some cases tolerance of exogenous methylases²⁴ can be enhanced by mutations in repair genes such as *Saccharomyces cerevisiae rad2*, or restriction genes such as *E. coli mcr* and *mrr*. We have tested five *E. coli* strains transfected with different foreign methylase genes. Application to eukaryotic genomes would seem to be feasible because methylase genes have been transfected into yeast and mammalian cells, where high levels of *in vivo* methylation of host DNA have been achieved without phenotypic consequences^{24,25}. With the approaching completion of several genome sequences, whole genome approaches like ours will become increasingly important and more efficient in defining biologically functional domains in the genome. □

Received 25 June; accepted 15 September 1992.

1. Razin, A. et al. *Nucleic Acids Res.* **8**, 1783-1792 (1980).
2. Geier, G. & Modrich, P. *J. Biol. Chem.* **254**, 1408-1413 (1979).
3. Blyn, L. B., Braaten, B. A. & Low, D. A. *EMBO J.* **9**, 4045-4054 (1990).

4. Ringquist, S. & Smith, C. L. *Proc. natn. Acad. Sci. U.S.A.* **89**, 4539-4543 (1992).
5. Braaten, B. A. et al. *Proc. natn. Acad. Sci. U.S.A.* **89**, 4250-4254 (1992).
6. Campbell, J. L. & Kleckner, N. *Gene* **74**, 189-190 (1988).
7. Jaworski, A. et al. *Science* **238**, 773-777 (1987).
8. Yang, C. C. & Nash, H. C. *Cell* **57**, 869-880 (1989).
9. Schultz, S. C., Shields, G. C. & Steitz, T. A. *Science* **253**, 1001-1007 (1991).
10. Wilson, G. G. *Gene* **74**, 281-289 (1988).
11. Valentin-Hansen, P. et al. *Molec. Microbiol.* **3**, 1385-1390 (1989).
12. Bird, A. P. & Southern, E. M. *J. molec. Biol.* **118**, 27-47 (1978).
13. Yamada, M. & Saier, M. J. *biol. Chem.* **262**, 5455-5462 (1987).
14. Lengeler, J. & Steinberger, H. *Molec. gen. Genet.* **164**, 163-170 (1978).
15. Sabourin, D. & Beckwith, J. *J. Bact.* **122**, 338-340 (1975).
16. Piette, J. et al. *Proc. natn. Acad. Sci. U.S.A.* **81**, 4134-4138 (1984).
17. Lyons, S. M. & Schendel, P. F. *J. Bact.* **159**, 421-423 (1984).
18. Campbell, J. L. & Kleckner, N. *Cell* **62**, 967-979 (1990).
19. Ephrussi, A., Church, G. M., Toneyawa, S. & Gilbert, W. *Science* **227**, 134-140 (1985).
20. Becker, M. M. & Wang, J. C. *Nature* **309**, 682-687 (1984).
21. Cartwright, I. & Kelly, S. E. *BioTechniques* **11**, 188-203 (1991).
22. Saluz, H. P., Wiebauer, K. & Wallace, A. *Trends Genet.* **7**, 207-211 (1991).
23. Singh, J. & Klar, A. J. S. *Genes Dev.* **6**, 186-196 (1992).
24. Feher, Z., Schlagman, S. L., Miner, Z. & Hattman, S. *Curr. Genet.* **16**, 461-464 (1989).
25. Kwok, T. J. et al. *Nucleic Acids Res.* **16**, 11489-11505 (1988).
26. Vovis, G. F. & Lacks, S. J. *J. molec. Biol.* **115**, 525-538 (1977).
27. Griboskov, M., Luthy, R. & Eisenberg, D. *Meth. Enzym.* **183**, 146-159 (1990).
28. Stormo, G. D. & Hartzell, G. W. *Proc. natn. Acad. Sci. U.S.A.* **86**, 1183-1187 (1989).
29. Jiang, W. et al. *Molec. Microbiol.* **4**, 2003-2006 (1990).
30. Davis, T., Yamada, M., Elgort, M. & Saier, M. H. *Molec. Microbiol.* **2**, 405-412 (1988).
31. Silverman, M. & Simon, M. *J. Bact.* **120**, 1196-1203 (1974).
32. Bartlett, B. H., Frantz, B. B. & Matsumura, P. *J. Bact.* **170**, 1575-1581 (1988).
33. Brissette, J. L., Weiner, L., Ripmaster, T. L. & Model, P. *Genbank* **69.0** (1991).
34. Shea, C. M. & McIntosh, M. A. *Genbank* **69.0** (1991).
35. Ebright, R. in *Molecular Structure and Biological Activity* (eds Griffen, J. & Duax, W.) (Elsevier Scientific, New York, 1982).
36. Gunasekera, A., Ebright, Y. W. & Ebright, R. H. *Nucleic Acids Res.* **18**, 6853-6856 (1990).
37. Church, G. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1991-1995 (1984).
38. Church, G. M. & Kieffer-Higgins, S. *Science* **240**, 185-188 (1988).

ACKNOWLEDGEMENTS. We thank R. Chin, G. Wilson, B. Bachmann, J. Roth, C. Smith, M. Rubenfield, A. Mian, R. Baldarelli and S. Kieffer-Higgins for strains, help and discussion. This work was supported by a grant from the Department of Energy. M.W. is the H. & C. Bower Fellow and is grateful to L. A. Donoso, W. S. Tasman, the Pennsylvania Lions Foundation, Research to Prevent Blindness, E. C. King Trust, and Crippled Children's Vitreo Retinal Research Foundation for support.

CORRECTIONS

The *Drosophila* clock gene *per* affects intercellular communication

T. A. Bargiello, L. Saez, M. K. Baylies, G. Gasic, M. W. Young & D. C. Spray

Nature **328**, 686-691 (1987)

EXPERIMENTS recently performed by one of the authors of this paper cast doubt on the involvement of *per* in gap junction-mediated intercellular communication. For details see Scientific Correspondence, page 542 of this issue.

Expression cloning of a human DNA repair gene involved in xeroderma pigmentosum group C

Randy Legerski & Carolyn Peterson

Nature **359**, 70-73 (1992)

THIS Letter in the 3 September issue contains an error affecting two codons in the sequence of the XPCC gene shown in Fig. 3a. The corrected sequence appears below, starting at position 612 of the cDNA.

GTG AAG TGG TTC ATT
V K W F I

This correction does not alter any of the conclusions of the paper and the open reading frame remains at 823 amino acids. The correct sequence has been submitted to the EMBL database. The accession number is X65024. We regret any inconvenience caused by this error.

GUIDE TO AUTHORS

PLEASE follow these guidelines so that your manuscript may be handled expeditiously.

Nature is an international journal covering all the sciences. Contributors should therefore bear in mind those readers who work in other fields and those for whom English is a second language, and write clearly and simply, avoiding unnecessary technical terminology. Space in the journal is limited, making competition for publication severe. Brevity is highly valued. One printed page of *Nature*, without interruptions, contains about 1,300 words.

Manuscripts are selected for publication according to editorial assessment of their suitability and reports from independent referees. They can be sent to London or Washington and should be addressed to the Editor. Manuscripts may be dealt with in either office, depending on the subject matter, and will where necessary be sent between offices by overnight courier. High priority cannot be given to pre-submission enquiries; in urgent cases they can be made in the form of a one-page fax. All manuscripts are acknowledged on receipt but fewer than half are sent for review. Those that are not reviewed are returned as rapidly as possible so that they may be submitted elsewhere without delay. Contributors may suggest reviewers; limited requests for the exclusion of specific reviewers are usually heeded. Manuscripts are usually sent to two or three reviewers, chosen for their expertise rather than their geographical location. Manuscripts accepted for publication are typeset from the London office.

Nature requests authors to deposit sequence and crystallographic data in the databases that exist for this purpose, and to mention availability of these data.

Once a manuscript is accepted for publication, contributors will receive proofs in about 4 weeks. *Nature's* staff will edit manuscripts with a view to brevity and clarity, so contributors should check proofs carefully. Manuscripts are generally published 2-3 weeks after receipt of corrected proofs. *Nature* does not exact page charges. Contributors receive a reprint order form with their proofs; reprint orders are processed after the manuscript is published and payment received.

Categories of paper

Review articles survey recent developments in a field. Most are commissioned but suggestions are welcome in the form of a one-page synopsis addressed to the Reviews Coordinator. Length is negotiable in advance.

Articles are research reports whose conclusions are of general interest and which are sufficiently rounded to be a substantial advance in understanding. They should not have more than 3,000 words of text (not including figure legends) or more than six display items and should not occupy more than five pages of *Nature*.

Articles start with a heading of 50-80 words written to advertise their content in general terms, to which editors will pay particular attention. The heading does not usually contain numbers, abbreviations or measurements. The introduction to the study is contained in the first two or three paragraphs of the article, which also briefly summarize its results and implications. Articles have fewer than 50 references and may contain a few short subheadings.

Letters are short reports of outstanding novel findings whose implications are general and important enough to be of interest to those outside the field. Letters should have 1,000 or fewer words of text and four or fewer display items. The first paragraph describes, in not more than 150 words and without the use of abbreviations, the background, rationale and chief conclusions of the study for the particular benefit of non-specialist readers. Letters do not have subheadings and should contain fewer than 30 references.

Commentary articles deal with issues in, or arising from, research that are also of interest to readers outside research. Some are commissioned but suggestions can be made to the Commentary Editor in the form of a one-page synopsis.

News and Views articles inform non-specialist readers about new scientific advances, sometimes in the form of a conference report. Most are commissioned but proposals can be made in advance to the News and Views Editor.

Scientific Correspondence is for discussion of topical scientific matters, including those published in *Nature*, and for miscellaneous contributions. Priority is given to letters of fewer than 500 words.

Preparation of manuscripts

All manuscripts should be typed, double-spaced, on one side of the paper only. An original and four copies are required, each accompanied by artwork. If photographs are included, five sets of originals are required; for line drawings, one set of originals and four good-quality photocopies are acceptable. Reference lists, figure legends and tables should all be on separate sheets, all of which should be double-spaced and numbered. Three copies of relevant manuscripts in press or submitted for publication elsewhere should be included with each copy of a submitted manuscript, and clearly marked as such. Five copies of revised and resubmitted manuscripts, labelled with their manuscript numbers are required, together with five copies of a letter detailing the changes made.

Titles are brief and simple. Active verbs, numerical values, abbreviations and punctuation are to be avoided. Titles should contain one or two key words for indexing purposes.

Artwork should be marked individually and clearly with the author's name and, when known, the manuscript number. Ideally, no figure should be larger than 28 by 22 cm. Figures with several parts are to be avoided and are permitted only if the parts are closely related, either experimentally or logically. Unlettered originals of photographs should be provided. Suggestions for cover illustrations, with captions and labelled with the manuscript number, are welcome. Original artwork is returned when a manuscript cannot be published.

Protein/nucleotide sequences should ideally be in the three-letter and not the single-letter code for amino acids. One column width of *Nature* can accommodate 20 amino acids or 60 base pairs.

Colour artwork. A charge of £500 per page is made as a contribution towards the cost of reproducing colour figures. Inability to pay these costs will not prevent publication of essential colour figures if the circumstances are explained. Proofs of colour artwork may be sent to contributors under separate cover from their galley proofs.

Figure legends should not exceed 300 words and ideally should be shorter. The figure is described first, then, briefly, the method. Reference to a method published elsewhere is preferable to a full description. Methods are not described in the text.

References are numbered sequentially as they appear in the text, followed by those in tables and finally by those in figure legends. Only papers published or in the press are numbered and included in the reference list. All other forms of reference should be cited in the text as a personal communication, manuscript submitted or in preparation. Text is not included in reference lists. References are abbreviated according to the *World List of Scientific Periodicals* (Butterworths, London, 1963-65). The first and last page numbers are included; reference to books should include publisher, place and date.

Abbreviations, symbols, units and Greek letters should be identified the first time they are used. Acronyms should be avoided whenever possible and, if used, defined. Footnotes are not used in the text.

Acknowledgements are brief and appear after the reference list; grant and contribution numbers are not allowed.

Supplementary information is material relevant to Articles or Letters which cannot, for lack of space, be published in full, but which is available from *Nature* on request.

Submission. Manuscripts can be sent to the Editor at 4 Little Essex Street, London WC2R 3LF, UK or at 1234 National Press Building, Washington, DC 20045, USA. Manuscripts or proofs sent by air courier to London should be declared as 'manuscripts' and 'value \$5' to prevent the imposition of import duty and value-added tax.

Scandinavian sourcebook

Products featured this week from companies based in Scandinavia include an RNA/DNA calculator, an automatic peptide synthesizer optimized for Fmoc polyamide chemistry and an ATP biomass kit.

To help molecular biologists monitor the progress of essential steps in mapping, cloning and sequencing, Pharmacia has introduced a quantification tool called the GeneQuant RNA/DNA calculator (*Reader Service No. 100*). GeneQuant uses diode-array technology for rapid multi-wavelength



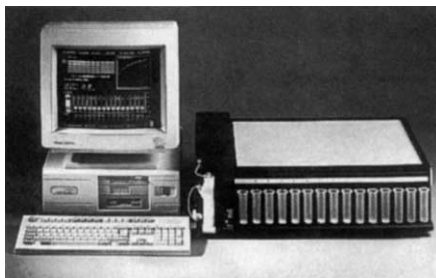
GeneQuant — Pharmacia's quick-check device for molecular biologists.

measurement and has no moving parts to provide increased reliability. The reverse-optics configuration also provides the convenience of an open sample compartment. The system requires only a few microlitres of your precious RNA/DNA sample and uses a disposable sample cell, which offers contamination-free sample recovery. The system also accommodates standard UV cells if sufficient sample volume is available. For convenience, GeneQuant uses preprogrammed formulae and all user-defined parameters are stored automatically. Features include (A_{260}/A_{280}) purity check and automatic quantification of RNA/DNA and oligonucleotides. Readings can be displayed in units of absorbance, weight and molar fraction. GeneQuant also calculates melting temperature and molarity for oligonucleotides.

Pharmacia has launched the **recombinant phage antibody system**, which is designed for researchers requiring large quantities of antibodies for experimentation or for use as reagents (*Reader Service No. 101*). The system, which was developed in collaboration with Cambridge Antibody Technology, is designed for cloning, expressing and detecting recombinant phage antibodies derived from a mouse B-cell source. Producing antibodies in *Escherichia coli* avoids the problems and expense of creating and maintaining hybridomas (which often prove to be unstable) in large-scale cell culture or in animal models. The system uses a 'phage-display' mode of expression in which the antibody fragment is displayed as a fusion with gene III protein

on the tip of M13 bacteriophage. When recombinant phage antibodies are reacted with an antigen of interest, phage display allows for affinity purification of specific phage and concomitant selection of antibody genes. Pharmacia says that this has proven useful for studies on the relationship between gene structure and antibody function, and for creating a library of antibodies with various antigen affinities. The system consists of a mouse ScFv module, an expression module and a detection module.

The **MK-III peptide synthesizer** from Kem-En-Tec is a fully automated, computer-controlled machine that is designed for the solid-phase synthesis of peptides in the 10–1,000-mg scale (*Reader Service No. 102*). To compensate for compression due to swelling of the resin during synthesis, the MK-III reactor is equipped with a stirring motor. The motor is activated briefly after injection of the reagent into the reactor. Kem-En-Tec says that this minimizes reagent consumption and ensures complete and uniform mixing of the reagent and the resin. The software features a real-time flow chart and a graphical display of conductiv-

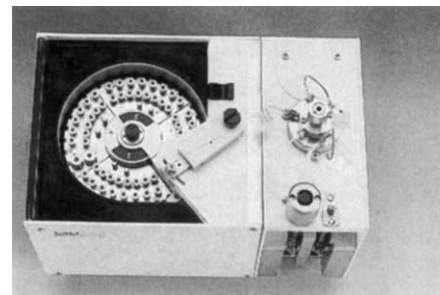


Solid-phase peptide synthesis is a snap with the MK-III from Kem-En-Tec.

ity monitoring on a PC. Peptides up to a length of 50 amino acids can be preprogrammed with individual parameters for the amino acid reagent in each step. The reagent rack is continuously updated so that the operator can leave the machine with 15 amino acids ready for automatic coupling. The protocols store information about reagent types, preactivation time, molecular weights, default reaction times and reagent excess for up to 5,760 amino acid derivatives. A special feature of the MK-III is the real-time monitoring system, which is based on automatic measurement of electrical conductivity in the reactor. This set-up optimizes solvent consumption and offers automatic feedback monitoring of the coupling steps during reactions involving preformed esters and symmetrical anhydrides. MK-III software runs on any IBM PC or compatible.

The Wizard **automatic benchtop gamma counter** from Wallac uses well-type detectors with almost 4π geometry (*Reader Service No. 103*). The counter features a solid 75-mm all-around lead shield as an effective method of cutting out background and crosstalk. The manufacturers of the new counter guarantee that crosstalk for ^{51}Cr is undetectable for any sample on the conveyor; for ^{59}Fe , the guaranteed figure is less than 0.04 per cent. The Wizard measures samples up to 2,000 keV. Vials can be of any size and shape up to 28 mm in diameter and up to 100 mm in length. The counter does not require the use of adapters when measuring tubes of variable shape, such as microcentrifuge or Eppendorf tubes. It has a capacity of up to 1,000 13-mm vials. Wallac says that for different sample volumes it is not necessary to carry out separate normalization routines. (Volume dependency is less than 1 per cent per ml for ^{59}Fe using a 20-ml vial.) The counter can be run using its own built-in software or from a PC. For data processing and quality control, Wallac offers MultiCalc PC software.

The **CMA/200 refrigerated micro-sampler** from CMA/Microdialysis is designed for reproducible submicrolitre injections of precious biological samples to be assayed by column liquid chromatography (*Reader Service No. 104*). The instrument has a minimum injection volume of 0.1 μl and only 0.5 μl of extra sample volume in the vial is required. The company says that each instrument has a relative standard deviation of less than 0.2 per cent for typical 10- μl injections. The CMA/200 has a vial capacity of either $64 \times 300\text{-}\mu\text{l}$ vials or $40 \times 2\text{-ml}$ vials and injection volumes can range from 0.1 μl to 1.2 ml. Programmable thermoelectric cooling is also offered, which is designed to prevent evaporation of samples and chemical degradation of labile molecules. The CMA/200 can now be controlled from either a Macintosh computer or a PC running DOS, Windows 3.1 or OS/2.



CMA/200 refrigerated micro-sampler provides temperature control down to 4 °C.

PRODUCT REVIEW

The expanded software package enables the computer to control up to two independent CMA/200 autosamplers, and at the same time run other programs for peak integration, statistics and word processing.

Heto offers a range of complete **vacuum concentration systems** for quick drying of liquid or frozen samples (*Reader Service No. 105*). The systems can be used to evaporate organic solvents as in HPLC fractions, dry acidic hydrolysates of proteins for amino acid analysis and concentrate clinical specimens for clinical diagnostic testing. A Heto system comprises a vacuum concentrator, a cooling trap with condenser temperature down to -65°C or -118°C , a vacuum pump and a mobile rack for integration of all components. In addition to the six complete systems that are available, Heto offers kits, freeze drying chambers for stoppering or bulk drying and manifolds for flask and ampoule drying, and a wide range of rotors taking sample volumes of up to 800 ml.

The new Sigma 70 from KSV Instruments facilitates fully computer-controlled and **automatic measurement of surface/interfacial tensions** and contact angles of liquids (*Reader Service No. 106*). The instrument uses the Wilhelmy plate or du Nouy ring method for force measurement and its super-sensitive fibre range balance operates at 0.25 mN to 25 mN range. Standard programmes include automatic determination of critical micelle concentration, push and pull surface and interfacial tension measurements, and measurement of dynamic advancing and receding contact angles. Measuring parameters are displayed online in clear digits on the interface unit and graphical result presentation can be monitored at the computer screen. All experimental data are stored on the hard disk and the flexible multitasking software allows customized application development.

The Micro96 **harvester** from Skatron Instruments is supplied in three different models, allowing users to harvest all 96 wells simultaneously (*Reader Service No.*

107). The harvesters feature more flexible programming to satisfy the most demanding requirements for improved quality of harvesting, including cell proliferation and receptor binding assays. The three models include the standard 6×16 format, the 8×12 microplate format and the macrowell version for tubes, with a standard 6×16 format.

The **bioluminescent assay of ATP** is now widely used for the rapid measurement of biomass in cell cultures. The reaction, which is catalysed by firefly luciferase, results in the formation of light which is then quantified by a simple luminometer. The resultant light intensity is proportional to the number of cells. The ATP Biomass Kit from Bio-Orbit uses patented continuous light output reagents, which allow the detection of cell numbers to be performed in



Bio-Orbit's ATP biomass kit.

less than 5 min (*Reader Service No. 108*). The kit is suitable for the determination of growth rate or inhibitory effects (biocide efficacy) in cell culture. When combined with a short pre-enrichment step, the kit can be used to reduce the time taken for sterility tests of both media and finished products.

Biohit has extended its range of Proline **electronic pipettes** with two new models for micro- and macro-volume pipetting (*Reader Service No. 109*). These units provide the user with four separate programs for pipetting, repeat dispensing, reverse pipetting and dilute-dispensing. In addition, a mixing protocol is provided in the pipetting and dilute-dispense mode of both pipettes. The micro-volume pipette covers the range from 0.2–10 μl and is intended for precise low-volume pipetting regimes like those used in molecular biology. It can be used with micro-volume pipette tips containing a filter. Large-volume pipetting regimes can be handled using the macro-volume pipette, which has a pipetting range of 100–5,000 μl .

The new Finnpiptette Digital ACL **autoclavable pipette** from Labsystems is designed to provide operators with increased protection against infection while working with biohazardous materials (*Reader Service No. 110*). It also safeguards the integrity of experiments by eliminating cross contamination. The pipette is de-tipped at the end of an experiment and then placed in the autoclave for sterilization at 121°C . In addition to being autoclavable, the pipette has a more rounded body to minimize hand

strain, a re-designed tip ejector and a positive click-turn mechanism for rapid and secure volume adjustment. Six models cover the range 0.5 μl to 10 ml.

The new GeNunc Modules from Nunc offer an **alternative to microcentrifuge tubes** for use in sequencing, restriction analyses, transformations and ligation reactions (*Reader Service No. 111*). Made of polypropylene, the GeNunc Modules (250, 100 or 25 μl per well formats) are strips of wells held in a frame of the standard 96-MicroWell size and format, and are supplied with a special sealing tape. Nunc says that the strips in the GeNunc Modules are DNase- and RNase-free.

Eka Nobel has introduced a high-performance aminophase 100 Å (pore-size) **spherical silica** — Kromasil KR100-HN₂ (*Reader Service No. 112*). Kromasil KR100-HN₂ is manufactured using monofunctional reagent giving a monolayer on the silica surface, and is subsequently end-capped. Eka Nobel says that KR100-HN₂ provides good selectivity for carbohydrates, has a high loading capacity and good chemical stability. It is available as bulk material in 5-, 7-, 10-, 13- and 16- μm particle sizes, or packed in high-pressure, slurry-packed columns.

Medi-Cult offers a full line of **serum-free media products for assisted fertilization**, all of which are based on a modification of Earle's balanced salt solution and Medi-Cult's synthetic serum replacement (SSR2) (*Reader Service No. 113*). They include: universal *in vitro* fertilization (IVF) medium, sperm preparation medium, flushing medium, freezing medium and Percoll medium. A new three days IVF M3 Medium is available, which is based on a modification of Ham's F10 and F12 plus SSR2 and HSA. Medi-Cult says that using its M3 Medium with replacement of maximum two embryos on day three results in an increased rate of deliveries, a reduced abortion rate and eliminates the occurrence of triplets.

Dynal has extended its range of superparamagnetic, monodispersed Dynabeads for **immunomagnetic cell separations**. Three new products are now available: Dynabeads M-450 rat anti-mouse IgG1, M-450 rat anti-mouse IgG2a and M-450 rat anti-mouse IgM (*Reader Service No. 114*). All products are provided in 2- or 10-ml vials with a concentration of 30 mg ml⁻¹. The antibodies used are monoclonal, subclass-specific and directed against the heavy chain of mouse immunoglobulin. Once the Dynabeads are coated, cell separation is designed to be fast, simple and reliable, yielding positively isolated cell subsets that are pure, viable and functionally active. □

For more information, fill in the reader service card bound inside the journal.

ADVERTISEMENT

\$3.50/Base
Custom DNA
Synthesis

No Set-up Fee,
Shipped in 24 hours.

1-800-788-4362
1-800-786-4362 Fax

KEYSTONE
Laboratories, Inc.

Reader Service No.33